HELMUT SCHWICHTENBERG

MATHEMATICAL LOGIC

-lecture notes-

Mathematical Logic

Helmut Schwichtenberg

Mathematisches Institut der Universität München
Wintersemester 2003/2004

Contents

CHAPTER 1

Logic

The main subject of Mathematical Logic is mathematical proof. In this introductory chapter we deal with the basics of formalizing such proofs. The system we pick for the representation of proofs is Gentzen's natural deduction, from [**8**]. Our reasons for this choice are twofold. First, as the name says this is a *natural* notion of formal proof, which means that the way proofs are represented corresponds very much to the way a careful mathematician writing out all details of an argument would go anyway. Second, formal proofs in natural deduction are closely related (via the so-called Curry-Howard correspondence) to terms in typed lambda calculus. This provides us not only with a compact notation for logical derivations (which otherwise tend to become somewhat unmanagable tree-like structures), but also opens up a route to applying the computational techniques which underpin lambda calculus.

Apart from classical logic we will also deal with more constructive logics: minimal and intuitionistic logic. This will reveal some interesting aspects of proofs, e.g. that it is possible und useful to distinguish beween existential proofs that actually construct witnessing objects, and others that don't. As an example, consider the following proposition.

There are irrational numbers a, b such that a^b is rational.

This can be proved as follows, by cases.

Case $\sqrt{2}^{\sqrt{2}}$ is rational. Choose $a = \sqrt{2}$ and $b = \sqrt{2}$. Then a, b are irrational and by assumption a^b is rational.

Case $\sqrt{2}^{\sqrt{2}}$ is irrational. Choose $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Then by assumption a, b are irrational and

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \left(\sqrt{2}\right)^2 = 2$$

is rational. □

As long as we have not decided whether $\sqrt{2}^{\sqrt{2}}$ is rational, we do not know which numbers a, b we must take. Hence we have an example of an existence proof which does not provide an instance.

An essential point for Mathematical Logic is to fix a formal language to be used. We take implication $\to$ and the universal quantifier $\forall$ as basic. Then the logic rules correspond to lambda calculus. The additional connectives $\bot$, $\exists$, $\vee$ and $\wedge$ are defined via axiom schemes. These axiom schemes will later be seen as special cases of introduction and elimination rules for inductive definitions.

1. Formal Languages

1.1. Terms and Formulas. Let a countable infinite set $\{ v_i \mid i \in \mathbb{N} \}$ of *variables* be given; they will be denoted by x, y, z. A first order language $\mathcal{L}$ then is determined by its *signature*, which is to mean the following.

- For every natural number $n \geq 0$ a (possible empty) set of n-ary *relation symbols* (also called *predicate symbols*). 0-ary relation symbols are called *propositional symbols*. $\bot$ (read "falsum") is required as a fixed propositional symbol. The language will *not*, unless stated otherwise, contain $=$ as a primitive.
- For every natural number $n \geq 0$ a (possible empty) set of n-ary *function symbols*. 0-ary function symbols are called *constants*.

We assume that all these sets of variables, relation and function symbols are disjoint.

For instance the language $\mathcal{L}_G$ of group theory is determined by the signature consisting of the following relation and function symbols: the group operation $\circ$ (a binary function symbol), the unit e (a constant), the inverse operation $^{-1}$ (a unary function symbol) and finally equality $=$ (a binary relation symbol).

$\mathcal{L}$-terms are inductively defined as follows.

- Every variable is an $\mathcal{L}$-term.
- Every constant of $\mathcal{L}$ is an $\mathcal{L}$-term.
- If $t_1, \ldots, t_n$ are $\mathcal{L}$-terms and f is an n-ary function symbol of $\mathcal{L}$ with $n \geq 1$, then $f(t_1, \ldots, t_n)$ is an $\mathcal{L}$-term.

From $\mathcal{L}$-terms one constructs *$\mathcal{L}$-prime formulas*, also called *atomic formulas* of $\mathcal{L}$: If $t_1, \ldots, t_n$ are terms and R is an n-ary relation symbol of $\mathcal{L}$, then $R(t_1, \ldots, t_n)$ is an $\mathcal{L}$-prime formula.

$\mathcal{L}$-formulas are inductively defined from $\mathcal{L}$-prime formulas by

- Every $\mathcal{L}$-prime formula is an $\mathcal{L}$-formula.
- If A and B are $\mathcal{L}$-formulas, then so are $(A \to B)$ ("if A, then B"), $(A \wedge B)$ ("A and B") and $(A \vee B)$ ("A or B").
- If A is an $\mathcal{L}$-formula and x is a variable, then $\forall x A$ ("for all x, A holds") and $\exists x A$ ("there is an x such that A") are $\mathcal{L}$-formulas.

Negation, classical disjunction, and the classical existential quantifier are defined by

$$\begin{aligned} \neg A &:= A \to \bot, \\ A \vee^{\mathsf{cl}} B &:= \neg A \to \neg B \to \bot, \\ \exists^{\mathsf{cl}} x A &:= \neg \forall x \neg A. \end{aligned}$$

Usually we fix a language $\mathcal{L}$, and speak of terms and formulas instead of $\mathcal{L}$-terms and $\mathcal{L}$-formulas. We use

r, s, t	for terms,
x, y, z	for variables,
c	for constants,
P, Q, R	for relation symbols,
f, g, h	for function symbols,
A, B, C, D	for formulas.

DEFINITION. The *depth* $\mathsf{dp}(A)$ of a formula A is the maximum length of a branch in its construction tree. In other words, we define recursively $\mathsf{dp}(P) = 0$ for atomic P, $\mathsf{dp}(A \circ B) = \max(\mathsf{dp}(A), \mathsf{dp}(B)) + 1$ for binary operators $\circ$, $\mathsf{dp}(\circ A) = \mathsf{dp}(A) + 1$ for unary operators $\circ$.

The *size* or *length* $|A|$ of a formula A is the number of occurrences of logical symbols and atomic formulas (parentheses not counted) in A: $|P| = 1$ for P atomic, $|A \circ B| = |A| + |B| + 1$ for binary operators $\circ$, $|\circ A| = |A| + 1$ for unary operators $\circ$.

One can show easily that $|A| + 1 \leq 2^{\mathsf{dp}(A)+1}$.

NOTATION (Saving on parentheses). In writing formulas we save on parentheses by assuming that $\forall, \exists, \neg$ bind more strongly than $\wedge, \vee$, and that in turn $\wedge, \vee$ bind more strongly than $\rightarrow, \leftrightarrow$ (where $A \leftrightarrow B$ abbreviates $(A \rightarrow B) \wedge (B \rightarrow A)$). Outermost parentheses are also usually dropped. Thus $A \wedge \neg B \rightarrow C$ is read as $((A \wedge (\neg B)) \rightarrow C)$. In the case of iterated implications we sometimes use the short notation

$$A_1 \rightarrow A_2 \rightarrow \dots A_{n-1} \rightarrow A_n \quad \text{for} \quad A_1 \rightarrow (A_2 \rightarrow \dots (A_{n-1} \rightarrow A_n) \dots).$$

To save parentheses in quantified formulas, we use a mild form of the *dot notation*: a dot immediately after $\forall x$ or $\exists x$ makes the scope of that quantifier as large as possible, given the parentheses around. So $\forall x.A \rightarrow B$ means $\forall x(A \rightarrow B)$, not $(\forall x A) \rightarrow B$.

We also save on parentheses by writing e.g. $Rxyz$, $Rt_0t_1t_2$ instead of $R(x, y, z)$, $R(t_0, t_1, t_2)$, where R is some predicate symbol. Similarly for a unary function symbol with a (typographically) simple argument, so fx for $f(x)$, etc. In this case no confusion will arise. But readability requires that we write in full $R(fx, gy, hz)$, instead of $Rfxgyhz$.

Binary function and relation symbols are usually written in *infix notation*, e.g. $x + y$ instead of $+(x, y)$, and $x < y$ instead of $<(x, y)$. We write $t \neq s$ for $\neg(t = s)$ and $t \not< s$ for $\neg(t < s)$.

1.2. Substitution, Free and Bound Variables. Expressions $\mathcal{E}, \mathcal{E}'$ which differ only in the names of bound variables will be regarded as identical. This is sometimes expressed by saying that $\mathcal{E}$ and $\mathcal{E}'$ are α-equivalent. In other words, we are only interested in expressions "modulo renaming of bound variables". There are methods of finding unique representatives for such expressions, for example the namefree terms of de Bruijn [**7**]. For the human reader such representations are less convenient, so we shall stick to the use of bound variables.

In the definition of "substitution of expression $\mathcal{E}'$ for variable x in expression $\mathcal{E}$", either one requires that *no* variable free in $\mathcal{E}'$ becomes bound by a variable-binding operator in $\mathcal{E}$, when the free occurrences of x are replaced by $\mathcal{E}'$ (also expressed by saying that there must be no "clashes of variables"), "$\mathcal{E}'$ *is free for* x *in* $\mathcal{E}$", or the substitution operation is taken to involve a systematic renaming operation for the bound variables, avoiding clashes. Having stated that we are only interested in expressions modulo renaming bound variables, we can without loss of generality assume that substitution is always possible.

Also, it is never a real restriction to assume that distinct quantifier occurrences are followed by distinct variables, and that the sets of bound and free variables of a formula are disjoint.

NOTATION. "FV" is used for the (set of) free variables of an expression; so $\mathsf{FV}(t)$ is the set of variables free in the term t, $\mathsf{FV}(A)$ the set of variables free in formula A etc.

$\mathcal{E}[x := t]$ denotes the result of substituting the term t for the variable x in the expression $\mathcal{E}$. Similarly, $\mathcal{E}[\vec{x} := \vec{t}]$ is the result of *simultaneously* substituting the terms $\vec{t} = t_1, \ldots, t_n$ for the variables $\vec{x} = x_1, \ldots, x_n$, respectively.

Locally we shall adopt the following convention. In an argument, once a formula has been introduced as $A(x)$, i.e., A with a designated variable x, we write $A(t)$ for $A[x := t]$, and similarly with more variables. □

1.3. Subformulas. Unless stated otherwise, the notion of *subformula* we use will be that of a subformula in the sense of Gentzen.

DEFINITION. (Gentzen) subformulas of A are defined by

(a) A is a subformula of A;
(b) if $B \circ C$ is a subformula of A then so are B, C, for $\circ = \rightarrow, \wedge, \vee$;
(c) if $\forall x B$ or $\exists x B$ is a subformula of A, then so is $B[x := t]$, for all t free for x in B.

If we replace the third clause by:

(c)′ if $\forall x B$ or $\exists x B$ is a subformula of A then so is B,

we obtain the notion of *literal* subformula.

DEFINITION. The notions of *positive, negative, strictly positive* subformula are defined in a similar style:

(a) A is a positive and a stricly positive subformula of itself;
(b) if $B \wedge C$ or $B \vee C$ is a positive [negative, strictly positive] subformula of A, then so are B, C;
(c) if $\forall x B$ or $\exists x B$ is a positive [negative, strictly positive] subformula of A, then so is $B[x := t]$;
(d) if $B \rightarrow C$ is a positive [negative] subformula of A, then B is a negative [positive] subformula of A, and C is a positive [negative] subformula of A;
(e) if $B \rightarrow C$ is a strictly positive subformula of A, then so is C.

A strictly positive subformula of A is also called a *strictly positive part (s.p.p.)* of A. Note that the set of subformulas of A is the union of the positive and negative subformulas of A. *Literal* positive, negative, strictly positive subformulas may be defined in the obvious way by restricting the clause for quantifiers.

EXAMPLE. $(P \rightarrow Q) \rightarrow R \wedge \forall x R'(x)$ has as s.p.p.'s the whole formula, $R \wedge \forall x R'(x)$, R, $\forall x R'(x)$, $R'(t)$. The positive subformulas are the s.p.p.'s and in addition P; the negative subformulas are $P \rightarrow Q$, Q.

2. Natural Deduction

We introduce Gentzen's system of natural deduction. To allow a direct correspondence with the lambda calculus, we restrict the rules used to those

for the logical connective $\to$ and the universal quantifier $\forall$. The rules come in pairs: we have an introduction and an elimination rule for each of these. The other logical connectives are introduced by means of axiom schemes: this is done for conjunction $\wedge$, disjunction $\vee$ and the existential quantifier $\exists$. The resulting system is called *minimal logic*; it has been introduced by Johannson in 1937 [**14**]. Notice that no negation is present.

If we then go on and require the *ex-falso-quodlibet* scheme for the nullary propositional symbol $\bot$ ("falsum"), we can embed *intuitionistic logic*. To obtain classical logic, we add as an axiom scheme the principle of *indirect proof*, also called *stability*. However, to obtain classical logic it suffices to restrict to the language based on $\to$, $\forall$, $\bot$ and $\wedge$; we can introduce classical disjunction $\vee^{\mathsf{cl}}$ and the classical existential quantifier $\exists^{\mathsf{cl}}$ via their (classical) definitions above. For these the usual introduction and elimination properties can then be derived.

2.1. Examples of Derivations. Let us start with some examples for natural proofs. Assume that a first order language $\mathcal{L}$ is given. For simplicity we only consider here proofs in pure logic, i.e. without assumptions (axioms) on the functions and relations used.

$$(A \to B \to C) \to (A \to B) \to A \to C \tag{1}$$

Assume $A \to B \to C$. To show: $(A \to B) \to A \to C$. So assume $A \to B$. To show: $A \to C$. So finally assume A. To show: C. We have A, by the last assumption. Hence also $B \to C$, by the first assumption, and B, using the next to last assumption. From $B \to C$ and B we obtain C, as required. □

$$(\forall x.A \to B) \to A \to \forall x B, \quad \text{if } x \notin \mathsf{FV}(A). \tag{2}$$

Assume $\forall x.A \to B$. To show: $A \to \forall x B$. So assume A. To show: $\forall x B$. Let x be arbitrary; note that we have not made any assumptions on x. To show: B. We have $A \to B$, by the first assumption. Hence also B, by the second assumption. □

$$(A \to \forall x B) \to \forall x.A \to B, \quad \text{if } x \notin \mathsf{FV}(A). \tag{3}$$

Assume $A \to \forall x B$. To show: $\forall x.A \to B$. Let x be arbitrary; note that we have not made any assumptions on x. To show: $A \to B$. So assume A. To show: B. We have $\forall x B$, by the first and second assumption. Hence also B. □

A characteristic feature of these proofs is that assumptions are introduced and eliminated again. At any point in time during the proof the free or "open" assumptions are known, but as the proof progresses, free assumptions may become cancelled or "closed" because of the implies-introduction rule.

We now reserve the word *proof* for the informal level; a formal representation of a proof will be called a *derivation*.

An intuitive way to communicate derivations is to view them as labelled trees. The labels of the inner nodes are the formulas derived at those points, and the labels of the leaves are formulas or terms. The labels of the nodes immediately above a node ν are the *premises* of the rule application, the formula at node ν is its *conclusion*. At the root of the tree we have the conclusion of the whole derivation. In natural deduction systems one works

with *assumptions* affixed to some leaves of the tree; they can be *open* or else *closed*.

Any of these assumptions carries a *marker*. As markers we use *assumption variables* $\Box_0, \Box_1, \ldots$, denoted by $u, v, w, u_0, u_1, \ldots$. The (previous) variables will now often be called *object variables*, to distinguish them from assumption variables. If at a later stage (i.e. at a node below an assumption) the dependency on this assumption is removed, we record this by writing down the assumption variable. Since the same assumption can be used many times (this was the case in example (1)), the assumption marked with u (and communicated by $u\colon A$) may appear many times. However, we insist that distinct assumption formulas must have distinct markers.

An inner node of the tree is understood as the result of passing form premises to a *conclusion*, as described by a given *rule*. The label of the node then contains in addition to the conclusion also the name of the rule. In some cases the rule binds or closes an assumption variable u (and hence removes the dependency of all assumptions $u\colon A$ marked with that u). An application of the $\forall$-introduction rule similarly binds an object variable x (and hence removes the dependency on x). In both cases the bound assumption or object variable is added to the label of the node.

2.2. Introduction and Elimination Rules for $\to$ and $\forall$. We now formulate the rules of natural deduction. First we have an assumption rule, that allows an arbitrary formula A to be put down, together with a marker u:

$$u\colon A \quad \text{Assumption}$$

The other rules of natural deduction split into introduction rules (I-rules for short) and elimination rules (E-rules) for the logical connectives $\to$ and $\forall$. For implication $\to$ there is an introduction rule $\to^+ u$ and an elimination rule $\to^-$, also called *modus ponens*. The left premise $A \to B$ in $\to^-$ is called *major premise* (or *main* premise), and the right premise A *minor premise* (or *side* premise). Note that with an application of the $\to^+ u$-rule *all* assumptions above it marked with $u\colon A$ are cancelled.

$$\frac{\begin{array}{c}[u\colon A]\\ \mid M\\ B\end{array}}{A \to B}\to^+ u \qquad \frac{\begin{array}{cc}\mid M & \mid N\\ A \to B & A\end{array}}{B}\to^-$$

For the universal quantifier $\forall$ there is an introduction rule $\forall^+ x$ and an elimination rule $\forall^-$, whose right premise is the term r to be substituted. The rule $\forall^+ x$ is subject to the following *(Eigen-) variable condition*: The derivation M of the premise A should not contain any open assumption with x as a free variable.

$$\frac{\begin{array}{c}\mid M\\ A\end{array}}{\forall x A}\forall^+ x \qquad \frac{\begin{array}{cc}\mid M & \\ \forall x A & r\end{array}}{A[x := r]}\forall^-$$

We now give derivations for the example formulas (1) – (3). Since in many cases the rule used is determined by the formula on the node, we

suppress in such cases the name of the rule,

$$\dfrac{\dfrac{\dfrac{u\colon A\to B\to C \qquad w\colon A}{B\to C} \qquad \dfrac{v\colon A\to B \qquad w\colon A}{B}}{C}\to^+ w}{A\to C}\to^+ v}{(A\to B)\to A\to C}\to^+ u \tag{1}$$

$$\dfrac{\dfrac{\dfrac{u\colon \forall x.A\to B \qquad x}{A\to B} \qquad v\colon A}{B}\forall^+ x}{\forall xB}\to^+ v}{A\to \forall xB}\to^+ u \tag{2}$$

Note here that the variable condition is satisfied: x is not free in A (and also not free in $\forall x.A\to B$).

$$\dfrac{\dfrac{\dfrac{u\colon A\to \forall xB \qquad v\colon A}{\forall xB} \qquad x}{B}\to^+ v}{A\to B}\forall^+ x}{\forall x.A\to B}\to^+ u \tag{3}$$

Here too the variable condition is satisfied: x is not free in A.

2.3. Axiom Schemes for Disjunction, Conjunction, Existence and Falsity. We follow the usual practice of considering all free variables in an axiom as universally quantified outside.

Disjunction. The introduction axioms are

$$\vee_0^+\colon A\to A\vee B$$
$$\vee_1^+\colon B\to A\vee B$$

and the elimination axiom is

$$\vee^-\colon (A\to C)\to (B\to C)\to A\vee B\to C.$$

Conjunction. The introduction axiom is

$$\wedge^+\colon A\to B\to A\wedge B$$

and the elimination axiom is

$$\wedge^-\colon (A\to B\to C)\to A\wedge B\to C.$$

Existential Quantifier. The introduction axiom is

$$\exists^+\colon A\to \exists xA$$

and the elimination axiom is

$$\exists^-\colon (\forall x.A\to B)\to \exists xA\to B \quad (x \text{ not free in } B).$$

Falsity. This example is somewhat extreme, since there is no introduction axiom; the elimination axiom is

$$\bot^- \colon \bot \to A.$$

In the literature this axiom is frequently called "ex-falso-quodlibet", written Efq. It clearly is derivable from its instances $\bot \to R\vec{x}$, for every relation symbol R.

Equality. The introduction axiom is

$$\mathsf{Eq}^+ \colon \mathsf{Eq}(x, x)$$

and the elimination axiom is

$$\mathsf{Eq}^- \colon \forall x R(x, x) \to \mathsf{Eq}(x, y) \to R(x, y).$$

It is an easy exercise to show that the usual equality axioms can be derived.

All these axioms can be seen as special cases of a general scheme, that of an *inductively defined predicate*, which is defined by some introduction rules and one elimination rule. We will study this kind of definition in full generality in Chapter 6. $\mathsf{Eq}(x, y)$ is a binary such predicate, $\bot$ is a nullary one, and $A \vee B$ another nullary one which however depends on the two parameter predicates A and B.

The desire to follow this general pattern is also the reason that we have chosen our rather strange $\wedge^-$-axiom, instead of the more obvious $A \wedge B \to A$ and $A \wedge B \to B$ (which clearly are equivalent).

2.4. Minimal, Intuitionistic and Classical Logic. We write $\vdash A$ and call A *derivable* (in *minimal logic*), if there is a derivation of A without free assumptions, from the axioms of 2.3 using the rules from 2.2, but *without using the ex-falso-quodlibet axiom, i.e., the elimination axiom $\bot^-$ for $\bot$*. A formula B is called derivable from assumptions $A_1, \dots, A_n$, if there is a derivation (without $\bot^-$) of B with free assumptions among $A_1, \dots, A_n$. Let Γ be a (finite or infinite) set of formulas. We write $\Gamma \vdash B$ if the formula B is derivable from finitely many assumptions $A_1, \dots, A_n \in \Gamma$.

Similarly we write $\vdash_i A$ and $\Gamma \vdash_i B$ if use of the ex-falso-quodlibet axiom is allowed; we then speak of derivability in *intuitionistic logic*.

For classical logic there is no need to use the full set of logical connectives: classical disjunction as well as the classical existential quantifier are defined, by $A \vee^{\mathsf{cl}} B := \neg A \to \neg B \to \bot$ and $\exists^{\mathsf{cl}} x A := \neg \forall x \neg A$. Moreover, when dealing with derivability we can even get rid of conjunction; this can be seen from the following lemma:

Lemma (Elimination of $\wedge$). *For each formula A built with the connectives $\to, \wedge, \forall$ there are formulas $A_1, \dots, A_n$ without $\wedge$ such that $\vdash A \leftrightarrow \bigwedge_{i=1}^n A_i$.*

Proof. Induction on A. **Case** $R\vec{t}$. Take $n = 1$ and $A_1 := R\vec{t}$. **Case** $A \wedge B$. By induction hypothesis, we have $A_1, \dots, A_n$ and $B_1, \dots, B_m$. Take $A_1, \dots, A_n, B_1, \dots, B_m$. **Case** $A \to B$. By induction hypothesis, we have $A_1, \dots, A_n$ and $B_1, \dots, B_m$. For the sake of notational simplicity assume $n = 2$ and $m = 3$. Then

$$\vdash (A_1 \wedge A_2 \to B_1 \wedge B_2 \wedge B_3)$$

$$\leftrightarrow (A_1 \to A_2 \to B_1) \land (A_1 \to A_2 \to B_2) \land (A_1 \to A_2 \to B_3).$$

Case $\forall x A$. By induction hypothesis for A, we have $A_1, \dots, A_n$. Take $\forall x A_1, \dots, \forall x A_n$, for

$$\vdash \forall x \bigwedge_{i=1}^{n} A_i \leftrightarrow \bigwedge_{i=1}^{n} \forall x A_i.$$

This concludes the proof. □

For the rest of this section, let us restrict to the language based on $\to$, $\forall$, $\bot$ and $\land$. We obtain *classical logic* by adding, for every relation symbol R distinct from $\bot$, the *principle of indirect proof* expressed as the so-called "stability axiom" (Stab_R):

$$\neg\neg R\vec{x} \to R\vec{x}.$$

Let

$$\mathsf{Stab} := \{\, \forall \vec{x}.\neg\neg R\vec{x} \to R\vec{x} \mid R \text{ relation symbol distinct from } \bot \,\}.$$

We call the formula A *classically derivable* and write $\vdash_c A$ if there is a derivation of A from stability assumptions Stab_R. Similarly we define classical derivability from Γ and write $\Gamma \vdash_c A$, i.e.

$$\Gamma \vdash_c A \quad :\Longleftrightarrow \quad \Gamma \cup \mathsf{Stab} \vdash A.$$

Theorem (Stability, or Principle of Indirect Proof). *For every formula A (of our language based on $\to$, $\forall$, $\bot$ and $\land$),*

$$\vdash_c \neg\neg A \to A.$$

Proof. Induction on A. For simplicity, in the derivation to be constructed we leave out applications of $\to^+$ at the end. **Case** $R\vec{t}$ with R distinct from $\bot$. Use Stab_R. **Case** $\bot$. Observe that $\neg\neg\bot \to \bot = ((\bot \to \bot) \to \bot) \to \bot$. The desired derivation is

$$\dfrac{v\colon (\bot \to \bot) \to \bot \qquad \dfrac{\dfrac{u\colon \bot}{\bot \to \bot} \to^+ u}{\bot}}{\bot}$$

Case $A \to B$. Use $\vdash (\neg\neg B \to B) \to \neg\neg(A \to B) \to A \to B$; a derivation is

$$\dfrac{u\colon \neg\neg B \to B \qquad \dfrac{\dfrac{v\colon \neg\neg(A \to B) \qquad \dfrac{\dfrac{u_1\colon \neg B \qquad \dfrac{u_2\colon A \to B \qquad w\colon A}{B}}{\bot}}{\neg(A \to B)} \to^+ u_2}{\bot}}{\neg\neg B} \to^+ u_1}{B}$$

Case $\forall x A$. Clearly it suffices to show $\vdash (\neg\neg A \to A) \to \neg\neg\forall x A \to A$; a derivation is

$$\dfrac{u\colon \neg\neg A \to A \qquad \dfrac{\dfrac{v\colon \neg\neg\forall x A \qquad \dfrac{\dfrac{u_1\colon \neg A \qquad \dfrac{u_2\colon \forall x A \qquad x}{A}}{\bot}}{\neg\forall x A} \to^+ u_2}{\bot}}{\neg\neg A} \to^+ u_1}{A}$$

The case $A \wedge B$ is left to the reader. $\square$

Notice that clearly $\vdash_c \bot \to A$, for stability is stronger:

$$\dfrac{\dfrac{\neg\neg A \to A \quad | \, M_{\mathsf{Stab}}}{A}}{\bot \to A} \to^+ u \qquad \dfrac{u\colon \bot}{\neg\neg A} \to^+ v^{\neg A}$$

where M_{Stab} is the (classical) derivation of stability.

Notice also that even for the $\to$, $\bot$-fragment the inclusion of minimal logic in intuitionistic logic, and of the latter in classical logic are proper. Examples are

$$\nvdash \bot \to P, \quad \text{but} \quad \vdash_i \bot \to P,$$
$$\nvdash_i ((P \to Q) \to P) \to P, \quad \text{but} \quad \vdash_c ((P \to Q) \to P) \to P.$$

Non-derivability can be proved by means of countermodels, using a semantic characterization of derivability; this will be done in Chapter 2. $\vdash_i \bot \to P$ is obvious, and the Peirce formula $((P \to Q) \to P) \to P$ can be derived in minimal logic from $\bot \to Q$ and $\neg\neg P \to P$, hence is derivable in classical logic.

2.5. Negative Translation. We embedd classical logic into minimal logic, via the so-called negative (or Gödel-Gentzen) translation.

A formula A is called *negative*, if every atomic formula of A distinct from $\bot$ occurs negated, and A does not contain $\vee$, $\exists$.

Lemma. *For negative A, $\vdash \neg\neg A \to A$.*

Proof. This follows from the proof of the stability theorem, using $\vdash \neg\neg\neg R\vec{t} \to \neg R\vec{t}$. $\square$

Since $\vee$, $\exists$ do not occur in formulas of classical logic, in the rest of this section we consider the language based on $\to$, $\forall$, $\bot$ and $\wedge$ only.

Definition (Negative translation ^{g} of Gödel-Gentzen).

$$\begin{aligned}
(R\vec{t})^g &:= \neg\neg R\vec{t} \qquad (R \text{ distinct from } \bot)\\
\bot^g &:= \bot,\\
(A \wedge B)^g &:= A^g \wedge B^g,\\
(A \to B)^g &:= A^g \to B^g,\\
(\forall x A)^g &:= \forall x A^g.
\end{aligned}$$

Theorem. *For all formulas A,*

(a) $\vdash_c A \leftrightarrow A^g$,
(b) $\Gamma \vdash_c A$ *iff* $\Gamma^g \vdash A^g$, *where* $\Gamma^g := \{\, B^g \mid B \in \Gamma \,\}$.

Proof. (a). The claim follows from the fact that $\vdash_c$ is compatible with equivalence. 2. $\Leftarrow$. Obvious $\Rightarrow$. By induction on the classical derivation. For a stability assumption $\neg\neg R\vec{t} \to R\vec{t}$ we have $(\neg\neg R\vec{t} \to R\vec{t})^g =$

$\neg\neg\neg\neg R\vec{t} \to \neg\neg R\vec{t}$, and this is easily derivable. *Case* $\to^+$. Assume

$$\dfrac{\begin{array}{c}[u\colon A]\\ \mathcal{D}\\ B\end{array}}{A \to B}\to^+ u$$

Then we have by induction hypothesis

$$\begin{array}{c}u\colon A^g\\ \mathcal{D}^g\\ B^g\end{array} \quad \text{hence} \quad \dfrac{\begin{array}{c}[u\colon A^g]\\ \mathcal{D}^g\\ B^g\end{array}}{A^g \to B^g}\to^+ u$$

Case $\to^-$. Assume

$$\dfrac{\begin{array}{cc}\mathcal{D}_0 & \mathcal{D}_1\\ A \to B & A\end{array}}{B}$$

Then we have by induction hypothesis

$$\begin{array}{cc}\mathcal{D}_0^g & \mathcal{D}_1^g\\ A^g \to B^g & A^g\end{array} \quad \text{hence} \quad \dfrac{\begin{array}{cc}\mathcal{D}_0^g & \mathcal{D}_1^g\\ A^g \to B^g & A^g\end{array}}{B^g}$$

The other cases are treated similarly. □

Corollary (Embedding of classical logic). *For negative A,*

$$\vdash_c A \iff \vdash A.$$

Proof. By the theorem we have $\vdash_c A$ iff $\vdash A^g$. Since A is negative, every atom distinct from $\bot$ in A must occur negated, and hence in A^g it must appear in threefold negated form (as $\neg\neg\neg R\vec{t}$). The claim follows from $\vdash \neg\neg\neg R\vec{t} \leftrightarrow \neg R\vec{t}$. □

Since every formula is classically equivalent to a negative formula, we have achieved an embedding of classical logic into minimal logic.

Note that $\nvdash \neg\neg P \to P$ (as we shall show in Chapter 2). The corollary therefore does not hold for all formulas A.

3. Normalization

We show in this section that every derivation can be transformed by appropriate conversion steps into a normal form. A derivation in normal form does not make "detours", or more precisely, it cannot occur that an elimination rule immediately follows an introduction rule. Derivations in normal form have many pleasant properties.

Uniqueness of normal form will be shown by means of an application of Newman's lemma; we will also introduce and discuss the related notions of confluence, weak confluence and the Church-Rosser property.

We finally show that the requirement to give a normal derivation of a derivable formula can sometimes be unrealistic. Following Statman [**25**] and Orevkov [**19**] we give examples of formulas C_k which are easily derivable with non-normal derivations (of size linear in k), but which require a non-elementary (in k) size in any normal derivation.

This can be seen as a theoretical explanation of the essential role played by lemmas in mathematical arguments.

3.1. Conversion. A conversion eliminates a detour in a derivation, i.e., an elimination immediately following an introduction. We consider the following conversions:

$\to$*-conversion.*

$$\dfrac{\dfrac{\begin{array}{c}[u\colon A]\\ \vert\ M\\ B\end{array}}{A \to B}\to^+ u \qquad \begin{array}{c}\vert\ N\\ A\end{array}}{B}\to^- \quad\mapsto\quad \begin{array}{c}\vert\ N\\ A\\ \vert\ M\\ B\end{array}$$

$\forall$*-conversion.*

$$\dfrac{\dfrac{\begin{array}{c}\vert\ M\\ A\end{array}}{\forall x A}\forall^+ x \qquad r}{A[x := r]}\forall^- \quad\mapsto\quad \begin{array}{c}\vert\ M'\\ A[x := r]\end{array}$$

3.2. Derivations as Terms. It will be convenient to represent derivations as terms, where the derived formula is viewed as the type of the term. This representation is known under the name *Curry-Howard correspondence*.

We give an inductive definition of derivation terms in the table below, where for clarity we have written the corresponding derivations to the left. For the universal quantifier $\forall$ there is an introduction rule $\forall^+ x$ and an elimination rule $\forall^-$, whose right premise is the term r to be substituted. The rule $\forall^+ x$ is subject to the following *(Eigen-) variable condition*: The derivation term M of the premise A should not contain any open assumption with x as a free variable.

3.3. Reduction, Normal Form. Although every derivation term carries a formula as its type, we shall usually leave these formulas implicit and write derivation terms without them.

Notice that every derivation term can be written uniquely in one of the forms

$$u\vec{M} \mid \lambda v M \mid (\lambda v M)N\vec{L},$$

where u is an assumption variable or assumption constant, v is an assumption variable or object variable, and M, N, L are derivation terms or object terms.

Here the final form is not normal: $(\lambda v M)N\vec{L}$ is called β*-redex* (for "reducible expression"). The *conversion rule* is

$$(\lambda v M)N \mapsto_\beta M[v := N].$$

Notice that in a substitution $M[v := N]$ with M a derivation term and v an object variable, one also needs to substitute in the formulas of M.

The *closure* of the conversion relation $\mapsto_\beta$ is defined by

- If $M \mapsto_\beta M'$, then $M \to M'$.
- If $M \to M'$, then also $MN \to M'N$, $NM \to NM'$, $\lambda v M \to \lambda v M'$ (*inner reductions*).

So $M \to N$ means that M *reduces in one step to* N, i.e., N is obtained from M by replacement of (an occurrence of) a redex M' of M by a conversum M'' of M', i.e. by a single conversion. The relation $\to^+$ ("*properly*

derivation	term
$u\colon A$	u^A
$\dfrac{\begin{array}{c}[u\colon A]\\ \mid M\\ B\end{array}}{A \to B} \to^+ u$	$(\lambda u^A M^B)^{A\to B}$
$\dfrac{\begin{array}{cc}\mid M & \mid N\\ A \to B & A\end{array}}{B} \to^-$	$(M^{A\to B} N^A)^B$
$\dfrac{\begin{array}{c}\mid M\\ A\end{array}}{\forall x A} \forall^+ x$ (with var.cond.)	$(\lambda x M^A)^{\forall x A}$ (with var.cond.)
$\dfrac{\begin{array}{cc}\mid M & \\ \forall x A & r\end{array}}{A[x := r]} \forall^-$	$(M^{\forall x A} r)^{A[x:=r]}$

TABLE 1. Derivation terms for $\to$ and $\forall$

reduces to") is the transitive closure of $\to$ and $\to^*$ ("*reduces to*") is the reflexive and transitive closure of $\to$. The relation $\to^*$ is said to be the notion of reduction *generated* by $\mapsto$. $\leftarrow$, $\leftarrow^+$, $\leftarrow^*$ are the relations converse to $\to, \to^+, \to^*$, respectively.

A term M is *in normal form*, or M is *normal*, if M does not contain a redex. M *has a normal form* if there is a normal N such that $M \to^* N$.

A *reduction sequence* is a (finite or infinite) sequence $M_0 \to M_1 \to M_2 \ldots$ such that $M_i \to M_{i+1}$, for all i.

Finite reduction sequences are partially ordered under the initial part relation; the collection of finite reduction sequences starting from a term M forms a tree, the *reduction tree* of M. The branches of this tree may be identified with the collection of all infinite and all terminating finite reduction sequences.

A term is *strongly normalizing* if its reduction tree is finite.

EXAMPLE.

$$(\lambda x \lambda y \lambda z.xz(yz))(\lambda u \lambda v\, u)(\lambda u' \lambda v'\, u') \to$$
$$(\lambda y \lambda z.(\lambda u \lambda v\, u)z(yz))(\lambda u' \lambda v'\, u') \quad \to$$

$$\begin{aligned}(\lambda y\lambda z.(\lambda v\, z)(yz))(\lambda u'\lambda v'\, u') &\to\\ (\lambda y\lambda z\, z)(\lambda u'\lambda v'\, u') &\to \lambda z\, z.\end{aligned}$$

LEMMA (Substitutivity of $\to$). (a) *If $M \to M'$, then $MN \to M'N$.*
(b) *If $N \to N'$, then $MN \to MN'$.*
(c) *If $M \to M'$, then $M[v := N] \to M'[v := N]$.*
(d) *If $N \to N'$, then $M[v := N] \to^* M[v := N']$.*

PROOF. (a) and (c) are proved by induction on $M \to M'$; (b) and (d) by induction on M. Notice that the reason for $\to^*$ in (d) is the fact that v may have many occurrences in M. □

3.4. Strong Normalization. We show that every term is strongly normalizing.

To this end, define by recursion on k a relation $\mathsf{sn}(M, k)$ between terms M and natural numbers k with the intention that k is an upper bound on the number of reduction steps up to normal form.

$$\begin{aligned}\mathsf{sn}(M, 0) &:\Longleftrightarrow M \text{ is in normal form,}\\ \mathsf{sn}(M, k+1) &:\Longleftrightarrow \mathsf{sn}(M', k) \text{ for all } M' \text{ such that } M \to M'.\end{aligned}$$

Clearly a term is strongly normalizable if there is a k such that $\mathsf{sn}(M, k)$. We first prove some closure properties of the relation sn.

LEMMA (Properties of sn). (a) *If $\mathsf{sn}(M, k)$, then $\mathsf{sn}(M, k+1)$.*
(b) *If $\mathsf{sn}(MN, k)$, then $\mathsf{sn}(M, k)$.*
(c) *If $\mathsf{sn}(M_i, k_i)$ for $i = 1 \dots n$, then $\mathsf{sn}(uM_1 \dots M_n, k_1 + \dots + k_n)$.*
(d) *If $\mathsf{sn}(M, k)$, then $\mathsf{sn}(\lambda vM, k)$.*
(e) *If $\mathsf{sn}(M[v := N]\vec{L}, k)$ and $\mathsf{sn}(N, l)$, then $\mathsf{sn}((\lambda vM)N\vec{L}, k + l + 1)$.*

PROOF. (a). Induction on k. Assume $\mathsf{sn}(M, k)$. We show $\mathsf{sn}(M, k+1)$. So let M' with $M \to M'$ be given; because of $\mathsf{sn}(M, k)$ we must have $k > 0$. We have to show $\mathsf{sn}(M', k)$. Because of $\mathsf{sn}(M, k)$ we have $\mathsf{sn}(M', k-1)$, hence by induction hypothesis $\mathsf{sn}(M', k)$.

(b). Induction on k. Assume $\mathsf{sn}(MN, k)$. We show $\mathsf{sn}(M, k)$. In case $k = 0$ the term MN is normal, hence also M is normal and therefore $\mathsf{sn}(M, 0)$. So let $k > 0$ and $M \to M'$; we have to show $\mathsf{sn}(M', k-1)$. From $M \to M'$ we have $MN \to M'N$. Because of $\mathsf{sn}(MN, k)$ we have by definition $\mathsf{sn}(M'N, k-1)$, hence $\mathsf{sn}(M', k-1)$ by induction hypothesis.

(c). Assume $\mathsf{sn}(M_i, k_i)$ for $i = 1 \dots n$. We show $\mathsf{sn}(uM_1 \dots M_n, k)$ with $k := k_1 + \dots + k_n$. Again we employ induction on k. In case $k = 0$ all M_i are normal, hence also $uM_1 \dots M_n$. So let $k > 0$ and $uM_1 \dots M_n \to M'$. Then $M' = uM_1 \dots M_i' \dots M_n$ with $M_i \to M_i'$; We have to show $\mathsf{sn}(uM_1 \dots M_i' \dots M_n, k-1)$. Because of $M_i \to M_i'$ and $\mathsf{sn}(M_i, k_i)$ we have $k_i > 0$ and $\mathsf{sn}(M_i', k_i - 1)$, hence $\mathsf{sn}(uM_1 \dots M_i' \dots M_n, k-1)$ by induction hypothesis.

(d). Assume $\mathsf{sn}(M, k)$. We have to show $\mathsf{sn}(\lambda vM, k)$. Use induction on k. In case $k = 0$ M is normal, hence λvM is normal, hence $\mathsf{sn}(\lambda vM, 0)$. So let $k > 0$ and $\lambda vM \to L$. Then L has the form $\lambda vM'$ with $M \to M'$. So $\mathsf{sn}(M', k-1)$ by definition, hence $\mathsf{sn}(\lambda vM', k)$ by induction hypothesis.

(e). Assume $\mathsf{sn}(M[v := N]\vec{L}, k)$ and $\mathsf{sn}(N, l)$. We need to show that $\mathsf{sn}((\lambda vM)N\vec{L}, k + l + 1)$. We use induction on $k + l$. In case $k + l = 0$ the

term N and $M[v := N]\vec{L}$ are normal, hence also M and all L_i. Hence there is exactly one term K such that $(\lambda vM)N\vec{L} \to K$, namely $M[v := N]\vec{L}$, and this K is normal. So let $k + l > 0$ and $(\lambda vM)N\vec{L} \to K$. We have to show $\mathsf{sn}(K, k + l)$.

Case $K = M[v := N]\vec{L}$, i.e. we have a head conversion. From $\mathsf{sn}(M[v := N]\vec{L}, k)$ we obtain $\mathsf{sn}(M[v := N]\vec{L}, k + l)$ by (a).

Case $K = (\lambda vM')N\vec{L}$ with $M \to M'$. Then we have $M[v := N]\vec{L} \to M'[v := N]\vec{L}$. Now $\mathsf{sn}(M[v := N]\vec{L}, k)$ implies $k > 0$ and $\mathsf{sn}(M'[v := N]\vec{L}, k - 1)$. The induction hypothesis yields $\mathsf{sn}((\lambda vM')N\vec{L}, k - 1 + l + 1)$.

Case $K = (\lambda vM)N'\vec{L}$ with $N \to N'$. Now $\mathsf{sn}(N, l)$ implies $l > 0$ and $\mathsf{sn}(N', l - 1)$. The induction hypothesis yields $\mathsf{sn}((\lambda vM)N'\vec{L}, k + l - 1 + 1)$, since $\mathsf{sn}(M[v := N']\vec{L}, k)$ by (a),

Case $K = (\lambda vM)N\vec{L}'$ with $L_i \to L_i'$ for some i and $L_j = L_j'$ for $j \neq i$. Then we have $M[v := N]\vec{L} \to M[v := N]\vec{L}'$. Now $\mathsf{sn}(M[v := N]\vec{L}, k)$ implies $k > 0$ and $\mathsf{sn}(M[v := N]\vec{L}', k - 1)$. The induction hypothesis yields $\mathsf{sn}((\lambda vM)N\vec{L}', k - 1 + l + 1)$. □

The essential idea of the strong normalization proof is to view the last three closure properties of sn from the preceding lemma without the information on the bounds as an inductive definition of a new set SN:

$$\frac{\vec{M} \in \mathsf{SN}}{u\vec{M} \in \mathsf{SN}}\,(\mathsf{Var}) \qquad \frac{M \in \mathsf{SN}}{\lambda vM \in \mathsf{SN}}\,(\lambda) \qquad \frac{M[v := N]\vec{L} \in \mathsf{SN} \qquad N \in \mathsf{SN}}{(\lambda vM)N\vec{L} \in \mathsf{SN}}\,(\beta)$$

Corollary. *For every term $M \in \mathsf{SN}$ there is a $k \in \mathbb{N}$ such that $\mathsf{sn}(M, k)$. Hence every term $M \in \mathsf{SN}$ is strongly normalizable*

Proof. By induction on $M \in \mathsf{SN}$, using the previous lemma. □

In what follows we shall show that *every* term is in SN and hence is strongly normalizable. Given the definition of SN we only have to show that SN is closed under application. In order to prove this we must prove simultaneously the closure of SN under substitution.

Theorem (Properties of SN). *For all formulas A, derivation terms $M \in \mathsf{SN}$ and $N^A \in \mathsf{SN}$ the following holds.*

(a) $M[v := N] \in \mathsf{SN}$.
(a') $M[x := r] \in \mathsf{SN}$.
(b) *Suppose M derives $A \to B$. Then $MN \in \mathsf{SN}$.*
(b') *Suppose M derives $\forall xA$. Then $Mr \in \mathsf{SN}$.*

Proof. By course-of-values induction on $\mathsf{dp}(A)$, with a side induction on $M \in \mathsf{SN}$. Let $N^A \in \mathsf{SN}$. We distinguish cases on the form of M.

Case $u\vec{M}$ by (Var) from $\vec{M} \in \mathsf{SN}$. (a). The SIH(a) (SIH means side induction hypothesis) yields $M_i[v := N] \in \mathsf{SN}$ for all M_i from $\vec{M}$. In case $u \neq v$ we immediately have $(u\vec{M})[v := N] \in \mathsf{SN}$. Otherwise we need $N\vec{M}[v := N] \in \mathsf{SN}$. But this follows by multiple applications of IH(b), since every $M_i[v := N]$ derives a subformula of A with smaller depth. (a'). Similar, and simpler. (b), (b'). Use (Var) again.

Case $\lambda v M$ by (λ) from $M \in \mathsf{SN}$. (a), (a'). Use (λ) again. (b). Our goal is $(\lambda v M)N \in \mathsf{SN}$. By (β) it suffices to show $M[v := N] \in \mathsf{SN}$ and $N \in \mathsf{SN}$. The latter holds by assumption, and the former by SIH(a). (b'). Similar, and simpler.

Case $(\lambda w M)K\vec{L}$ by (β) from $M[w := K]\vec{L} \in \mathsf{SN}$ and $K \in \mathsf{SN}$. (a). The SIH(a) yields $M[v := N][w := K[v := N]]\vec{L}[v := N] \in \mathsf{SN}$ and $K[v := N] \in \mathsf{SN}$, hence $(\lambda w M[v := N])K[v := N]\vec{L}[v := N] \in \mathsf{SN}$ by (β). (a'). Similar, and simpler. (b), (b'). Use (β) again. □

Corollary. *For every term we have $M \in \mathsf{SN}$; in particular every term M is strongly normalizable.*

Proof. Induction on the (first) inductive definition of derivation terms M. In cases u and $\lambda v M$ the claim follows from the definition of SN, and in case MN it follows from the preceding theorem. □

3.5. Confluence. A relation R is said to be *confluent*, or to have the *Church–Rosser property (CR)*, if, whenever $M_0 \, R \, M_1$ and $M_0 \, R \, M_2$, then there is an M_3 such that $M_1 \, R \, M_3$ and $M_2 \, R \, M_3$. A relation R is said to be *weakly confluent*, or to have the *weak Church–Rosser property (WCR)*, if, whenever $M_0 \, R \, M_1, M_0 \, R \, M_2$ then there is an M_3 such that $M_1 \, R^* \, M_3$ and $M_2 \, R^* \, M_3$, where R^* is the reflexive and transitive closure of R.

Clearly for a confluent reduction relation $\to^*$ the normal forms of terms are unique.

Lemma (Newman 1942). *Let $\to^*$ be the transitive and reflexive closure of $\to$, and let $\to$ be weakly confluent. Then the normal form w.r.t. $\to$ of a strongly normalizing M is unique. Moreover, if all terms are strongly normalizing w.r.t. $\to$, then the relation $\to^*$ is confluent.*

Proof. Call M *good* if it satisfies the confluence property w.r.t. $\to^*$, i.e. if whenever $K \leftarrow^* M \to^* L$, then $K \to^* N \leftarrow^* L$ for some N. We show that every strongly normalizing M is good, by transfinite induction on the well-founded partial order $\to^+$, restricted to all terms occurring in the reduction tree of M. So let M be given and assume

$$\forall M'.M \to^+ M' \Longrightarrow M' \text{ is good.}$$

We must show that M is good, so assume $K \leftarrow^* M \to^* L$. We may further assume that there are M', M'' such that $K \leftarrow^* M' \leftarrow M \to M'' \to^* L$, for otherwise the claim is trivial. But then the claim follows from the assumed weak confluence and the induction hypothesis for M' and M'', as shown in the picture below. □

3.6. Uniqueness of Normal Forms. We first show that $\to$ is weakly confluent. From this and the fact that it is strongly normalizing we can easily infer (using Newman's Lemma) that the normal forms are unique.

Proposition. *$\to$ is weakly confluent.*

Proof. Assume $N_0 \leftarrow M \to N_1$. We show that $N_0 \to^* N \leftarrow^* N_1$ for some N, by induction on M. If there are two inner reductions both on the same subterm, then the claim follows from the induction hypothesis using substitutivity. If they are on distinct subterms, then the subterms do not

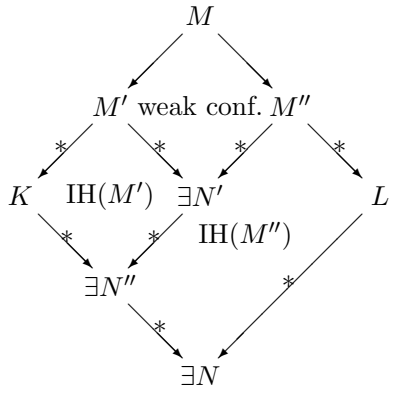
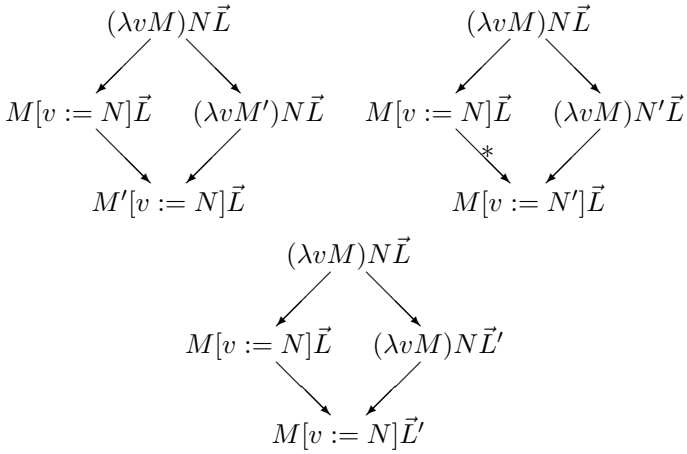


TABLE 2. Proof of Newman's lemma

overlap and the claim is obvious. It remains to deal with the case of a head reduction together with an inner conversion.



where for the lower left arrows we have used substitutivity again. □

COROLLARY. *Every term is strongly normalizing, hence normal forms are unique.* □

3.7. The Structure of Normal Derivations. Let M be a normal derivation, viewed as a prooftree. A sequence of f.o.'s (formula occurrences) $A_0, \dots, A_n$ such that (1) A_0 is a top formula (leaf) of the prooftree, and for $0 \leq i < n$, (2) A_{i+1} is immediately below A_i, and (3) A_i is not the minor premise of an $\to^-$-application, is called a *track* of the deduction tree M. A track *of order* 0 ends in the conclusion of M; a track of *order* $n+1$ ends in the minor premise of an $\to^-$-application with major premise belonging to a track of order n.

Since by normality an E-rule cannot have the conclusion of an I-rule as its major premise, the E-rules have to precede the I-rules in a track, so the following is obvious: a track may be divided into an E-part, say $A_0, \dots, A_{i-1}$, a minimal formula A_i, and an I-part $A_{i+1}, \dots, A_n$. In the E-part all rules

are E-rules; in the I-part all rules are I-rules; A_i is the conclusion of an E-rule and, if $i < n$, a premise of an I-rule. It is also easy to see that each f.o. of M belongs to some track. Tracks are pieces of branches of the tree with successive f.o.'s in the subformula relationship: either A_{i+1} is a subformula of A_i or vice versa. As a result, all formulas in a track $A_0, \ldots, A_n$ are subformulas of A_0 or of A_n; and from this, by induction on the order of tracks, we see that every formula in M is a subformula either of an open assumption or of the conclusion. To summarize, we have seen:

LEMMA. *In a normal derivation each formula occurrence belongs to some track.*

PROOF. By induction on the height of normal derivations. □

THEOREM. *In a normal derivation each formula is a subformula of either the end formula or else an assumption formula.*

PROOF. We prove this for tracks of order n, by induction on n. □

3.8. Normal Versus Non-Normal Derivations. We now show that the requirement to give a normal derivation of a derivable formula can sometimes be unrealistic. Following Statman [**25**] and Orevkov [**19**] we give examples of formulas C_k which are easily derivable with non-normal derivations (whose number of nodes is linear in k), but which require a non-elementary (in k) number of nodes in any normal derivation.

The example is related to Gentzen's proof in [**9**] of transfinite induction up to ω_k in arithmetic. There the function $y \oplus \omega^x$ plays a crucial role, and also the assignment of a "lifting"-formula $A^+(x)$ to any formula $A(x)$, by

$$A^+(x) := \forall y.(\forall z \prec y)A(z) \to (\forall z \prec y \oplus \omega^x)A(z).$$

Here we consider the numerical function $y + 2^x$ instead, and axiomatize its graph by means of Horn clauses. The formula C_k expresses that from these axioms the existence of 2_k follows. A short, non-normal proof of this fact can then be given by a modification of Gentzen's idea, and it is easily seen that any normal proof of C_k must contain at least 2_k nodes.

The derivations to be given make heavy use of the existential quantifier $\exists^{\mathsf{cl}}$ defined by $\neg\forall\neg$. In particular we need:

LEMMA (Existence Introduction). $\vdash A \to \exists^{\mathsf{cl}} x A$.

PROOF. $\lambda u^A \lambda v^{\forall x \neg A}.vxu$. □

LEMMA (Existence Elimination). $\vdash (\neg\neg B \to B) \to \exists^{\mathsf{cl}} x A \to (\forall x.A \to B) \to B$ *if* $x \notin FV(B)$.

PROOF. $\lambda u^{\neg\neg B \to B} \lambda v^{\neg\forall x \neg A} \lambda w^{\forall x.A \to B}.u\lambda u_2^{\neg B}.v\lambda x \lambda u_1^A.u_2(wxu_1)$. □

Note that the stability assumption $\neg\neg B \to B$ is not needed if B does not contain an atom $\neq \bot$ as a strictly positive subformula. This will be the case for the derivations below, where B will always be a classical existential formula.

Let us now fix our language. We use a ternary relation symbol R to represent the graph of the function $y + 2^x$; so $R(y, x, z)$ is intended to mean $y + 2^x = z$. We now axiomatize R by means of Horn clauses. For simplicity we use a unary function symbol s (to be viewed as the successor function)

and a constant 0; one could use logic without function symbols instead – as Orevkov does –, but this makes the formulas somewhat less readable and the proofs less perspicious. Note that Orevkov's result is an adaption of a result of Statman [**25**] for languages containing function symbols.

$$\mathsf{Hyp}_1 \colon \forall y R(y, 0, s(y))$$
$$\mathsf{Hyp}_2 \colon \forall y, x, z, z_1 . R(y, x, z) \to R(z, x, z_1) \to R(y, s(x), z_1)$$

The goal formula then is

$$C_k := \exists^{\mathsf{cl}} z_k, \ldots, z_0 . R(0, 0, z_k) \wedge R(0, z_k, z_{k-1}) \wedge \ldots \wedge R(0, z_1, z_0).$$

To obtain the short proof of the goal formula C_k we use formulas $A_i(x)$ with a free parameter x.

$$A_0(x) := \forall y \exists^{\mathsf{cl}} z \, R(y, x, z),$$
$$A_{i+1}(x) := \forall y . A_i(y) \to \exists^{\mathsf{cl}} z . A_i(z) \wedge R(y, x, z).$$

For the two lemmata to follow we give an informal argument, which can easily be converted into a formal proof. Note that the existence elimination lemma is used only with existential formulas as conclusions. Hence it is not necessary to use stability axioms and we have a derivation in minimal logic.

LEMMA. $\vdash \mathsf{Hyp}_1 \to \mathsf{Hyp}_2 \to A_i(0)$.

PROOF. **Case** $i = 0$. Obvious by Hyp_1.

Case $i = 1$. Let x with $A_0(x)$ be given. It is sufficient to show $A_0(s(x))$, that is $\forall y \exists^{\mathsf{cl}} z_1 R(y, s(x), z_1)$. So let y be given. We know

$$A_0(x) = \forall y \exists^{\mathsf{cl}} z \, R(y, x, z). \tag{4}$$

Applying (4) to our y gives z such that $R(y, x, z)$. Applying (4) again to this z gives z_1 such that $R(z, x, z_1)$. By Hyp_2 we obtain $R(y, s(x), z_1)$.

Case $i + 2$. Let x with $A_{i+1}(x)$ be given. It suffices to show $A_{i+1}(s(x))$, that is $\forall y . A_i(y) \to \exists^{\mathsf{cl}} z . A_i(z) \wedge R(y, s(x), z)$. So let y with $A_i(y)$ be given. We know

$$A_{i+1}(x) = \forall y . A_i(y) \to \exists^{\mathsf{cl}} z_1 . A_i(z_1) \wedge R(y, x, z_1). \tag{5}$$

Applying (5) to our y gives z such that $A_i(z)$ and $R(y, x, z)$. Applying (5) again to this z gives z_1 such that $A_i(z_1)$ and $R(z, x, z_1)$. By Hyp_2 we obtain $R(y, s(x), z_1)$. □

Note that the derivations given have a fixed length, independent of i.

LEMMA. $\vdash \mathsf{Hyp}_1 \to \mathsf{Hyp}_2 \to C_k$.

PROOF. $A_k(0)$ applied to 0 and $A_{k-1}(0)$ yields z_k with $A_{k-1}(z_k)$ such that $R(0, 0, z_k)$.

$A_{k-1}(z_k)$ applied to 0 and $A_{k-2}(0)$ yields z_{k-1} with $A_{k-2}(z_{k-1})$ such that $R(0, z_k, z_{k-1})$.

$A_1(z_2)$ applied to 0 and $A_0(0)$ yields z_1 with $A_0(z_1)$ such that $R(0, z_2, z_1)$.

$A_0(z_1)$ applied to 0 yields z_0 with $R(0, z_1, z_0)$. □

Note that the derivations given have length linear in k.

We want to compare the length of this derivation of C_k with the length of an arbitrary normal derivation.

PROPOSITION. *Any normal derivation of C_k from Hyp_1 and Hyp_2 has at least 2_k nodes.*

PROOF. Let a normal derivation M of falsity $\bot$ from Hyp_1, Hyp_2 and the additional hypothesis

$$u\colon \forall z_k,\dots,z_0.R(0,0,z_k)\to R(0,z_k,z_{k-1})\to\cdots\to R(0,z_1,z_0)\to\bot$$

be given. We may assume that M does not contain free object variables (otherwise substitute them by 0). The main branch of M must begin with u, and its side premises are all of the form $R(0,s^n(0),s^k(0))$.

Observe that any normal derivation of $R(s^m(0),s^n(0),s^k(0))$ from Hyp_1, Hyp_2 and u has at least 2^n occurrences of Hyp_1 and is such that $k=m+2^n$. This can be seen easily by induction on n. Note also that such a derivation cannot involve u.

If we apply this observation to the above derivations of the side premises we see that they derive

$$R(0,0,s^{2^0}(0)),\quad R(0,s^{2^0}(0),s^{2^{2^0}}(0)),\quad\dots\quad R(0,s^{2_{k-1}}(0),s^{2_k}(0)).$$

The last of these derivations uses at least $2^{2_{k-1}}=2_k$-times Hyp_1. □

4. Normalization including Permutative Conversions

The elimination of "detours" done in Section 3 will now be extended to the full language. However, incorporation of $\vee$, $\wedge$ and $\exists$ leads to difficulties. If we do this by means of axioms (or constant derivation terms, as in 2.3), we cannot read off as much as we want from a normal derivation. If we do it in the form of rules, we must also allow *permutative conversion*. The reason for the difficulty is that in the elimination rules for $\vee,\wedge,\exists$ the minor premise reappears in the conclusion. This gives rise to a situation where we first introduce a logical connective, then do not touch it (by carrying it along in minor premises of $\vee^-,\wedge^-,\exists^-$), and finally eliminate the connective. This is not a detour as we have treated them in Section 3, and the conversion introduced there cannot deal with this situation. What has to be done is a permutative conversion: permute an elimination immediately following an $\vee^-,\wedge^-,\exists^-$-rule over this rule to the minor premise.

We will show that any sequence of such conversion steps terminates in a normal form, which in fact is uniquely determined (again by Newman's lemma).

Derivations in normal form have many pleasant properties, for instance:

Subformula property: every formula occurring in a normal derivation is a subformula of either the conclusion or else an assumption;

Explicit definability: a normal derivation of a formula $\exists xA$ from assumptions not involving disjunctive of existential strictly positive parts ends with an existence introduction, hence also provides a term r and a derivation of $A[x:=r]$;

Disjunction property: a normal derivation of a disjunction $A\vee B$ from assumptions not involving disjunctions as strictly positive parts ends with a disjunction introduction, hence also provides either a derivation of A or else one of B;

4.1. Rules for $\vee$, $\wedge$ and $\exists$. Notice that we have not given rules for the connectives $\vee$, $\wedge$ and $\exists$. There are two reasons for this omission:

- They can be covered by means of appropriate axioms as constant derivation terms, as given in 2.3;
- For simplicity we want our derivation terms to be pure lambda terms formed just by lambda abstraction and application. This would be violated by the rules for $\vee$, $\wedge$ and $\exists$, which require additional constructs.

However – as just noted – in order to have a normalization theorem with a useful subformula property as a consequence we do need to consider rules for these connectives. So here they are:

Disjunction. The introduction rules are

$$\frac{\begin{array}{c}|\,M\\ A\end{array}}{A\vee B}\vee_0^+ \qquad \frac{\begin{array}{c}|\,M\\ B\end{array}}{A\vee B}\vee_1^+$$

and the elimination rule is

$$\frac{\begin{array}{ccc} & [u\colon A] & [v\colon B]\\ |\,M & |\,N & |\,K\\ A\vee B & C & C\end{array}}{C}\vee^- u,v$$

Conjunction. The introduction rule is

$$\frac{\begin{array}{cc}|\,M & |\,N\\ A & B\end{array}}{A\wedge B}\wedge^+$$

and the elimination rule is

$$\frac{\begin{array}{ccc} & [u\colon A] & [v\colon B]\\ |\,M & & |\,N\\ A\wedge B & & C\end{array}}{C}\wedge^- u,v$$

Existential Quantifier. The introduction rule is

$$\frac{\begin{array}{cc} & |\,M\\ r & A[x:=r]\end{array}}{\exists x A}\exists^+$$

and the elimination rule is

$$\frac{\begin{array}{cc} & [u\colon A]\\ |\,M & |\,N\\ \exists x A & B\end{array}}{B}\exists^- x,u \text{ (var.cond.)}$$

The rule $\exists^- x,u$ is subject to the following *(Eigen-) variable condition*: The derivation N should not contain any open assumptions apart from $u\colon A$ whose assumption formula contains x free, and moreover B should not contain the variable x free.

It is easy to see that for each of the connectives $\vee$, $\wedge$, $\exists$ the rules and the axioms are equivalent, in the sense that from the axioms and the premises of a rule we can derive its conclusion (of course without any $\vee,\wedge,\exists$-rules),

and conversely that we can derive the axioms by means of the $\vee, \wedge, \exists$-rules. This is left as an exercise.

The left premise in each of the elimination rules $\vee^-$, $\wedge^-$ and $\exists^-$ is called *major premise* (or *main* premise), and each of the right premises *minor premise* (or *side* premise).

4.2. Conversion. In addition to the $\to, \forall$-conversions treated in 3.1, we consider the following conversions:

$\vee$*-conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\, M\\ A\end{array}}{A \vee B}\vee_0^+ \qquad \begin{array}{c}[u\colon A]\\ |\, N\\ C\end{array} \qquad \begin{array}{c}[v\colon B]\\ |\, K\\ C\end{array}}{C}\vee^- u, v \quad \mapsto \quad \begin{array}{c}|\, M\\ A\\ |\, N\\ C\end{array}$$

and

$$\dfrac{\dfrac{\begin{array}{c}|\, M\\ B\end{array}}{A \vee B}\vee_1^+ \qquad \begin{array}{c}[u\colon A]\\ |\, N\\ C\end{array} \qquad \begin{array}{c}[v\colon B]\\ |\, K\\ C\end{array}}{C}\vee^- u, v \quad \mapsto \quad \begin{array}{c}|\, M\\ B\\ |\, K\\ C\end{array}$$

$\wedge$*-conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\, M\\ A\end{array} \qquad \begin{array}{c}|\, N\\ B\end{array}}{A \wedge B}\wedge^+ \qquad \begin{array}{c}[u\colon A]\ [v\colon B]\\ |\, K\\ C\end{array}}{C}\wedge^- u, v \quad \mapsto \quad \begin{array}{c}|\, M\\ A\end{array} \quad \begin{array}{c}|\, N\\ B\\ |\, K\\ C\end{array}$$

$\exists$*-conversion.*

$$\dfrac{\dfrac{r \qquad \begin{array}{c}|\, M\\ A[x := r]\end{array}}{\exists x A}\exists^+ \qquad \begin{array}{c}[u\colon A]\\ |\, N\\ B\end{array}}{B}\exists^- x, u \quad \mapsto \quad \begin{array}{c}|\, M\\ A[x := r]\\ |\, N'\\ B\end{array}$$

4.3. Permutative Conversion. In a permutative conversion we permute an E-rule upwards over the minor premises of $\vee^-$, $\wedge^-$ or $\exists^-$.

$\vee$*-perm conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\, M\\ A \vee B\end{array} \qquad \begin{array}{c}|\, N\\ C\end{array} \qquad \begin{array}{c}|\, K\\ C\end{array}}{C} \qquad \begin{array}{c}|\, L\\ C'\end{array}}{D}\text{E-rule} \quad \mapsto$$

$$\dfrac{\begin{array}{c}|\, M\\ A \vee B\end{array} \qquad \dfrac{\begin{array}{c}|\, N\\ C\end{array} \qquad \begin{array}{c}|\, L\\ C'\end{array}}{D}\text{E-rule} \qquad \dfrac{\begin{array}{c}|\, K\\ C\end{array} \qquad \begin{array}{c}|\, L\\ C'\end{array}}{D}\text{E-rule}}{D}$$

$\wedge$*-perm conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\,M\\ A\wedge B\end{array}\qquad\begin{array}{c}|\,N\\ C\end{array}}{C}\qquad\begin{array}{c}|\,K\\ C'\end{array}}{D}\ \text{E-rule}\quad\mapsto$$

$$\dfrac{\begin{array}{c}|\,M\\ A\wedge B\end{array}\qquad\dfrac{\begin{array}{c}|\,N\\ C\end{array}\qquad\begin{array}{c}|\,K\\ C'\end{array}}{D}\ \text{E-rule}}{D}$$

$\exists$*-perm conversion.*

$$\dfrac{\dfrac{\begin{array}{c}|\,M\\ \exists xA\end{array}\qquad\begin{array}{c}|\,N\\ B\end{array}}{B}\qquad\begin{array}{c}|\,K\\ C\end{array}}{D}\ \text{E-rule}\quad\mapsto$$

$$\dfrac{\begin{array}{c}|\,M\\ \exists xA\end{array}\qquad\dfrac{\begin{array}{c}|\,N\\ B\end{array}\qquad\begin{array}{c}|\,K\\ C\end{array}}{D}\ \text{E-rule}}{D}$$

4.4. Derivations as Terms. The term representation of derivations has to be extended. The rules for $\vee$, $\wedge$ and $\exists$ with the corresponding terms are given in the table below.

The introduction rule $\exists^+$ has as its left premise the witnessing term r to be substituted. The elimination rule $\exists^- u$ is subject to an *(Eigen-) variable condition*: The derivation term N should not contain any open assumptions apart from $u\colon A$ whose assumption formula contains x free, and moreover B should not contain the variable x free.

4.5. Permutative Conversions. In this section we shall write derivation terms without formula superscripts. We usually leave implicit the extra (formula) parts of derivation constants and for instance write $\exists^+$, $\exists^-$ instead of $\exists^+_{x,A}$, $\exists^-_{x,A,B}$. So we consider derivation terms M, N, K of the forms

$$u \mid \lambda v M \mid \lambda y M \mid \vee_0^+ M \mid \vee_1^+ M \mid \langle M, N\rangle \mid \exists^+ r M \mid$$
$$MN \mid Mr \mid M(v_0.N_0, v_1.N_1) \mid M(v, w.N) \mid M(v.N);$$

in these expressions the variables y, v, v_0, v_1, w get bound.

To simplify the technicalities, we restrict our treatment to the rules for $\to$ and $\exists$. It can easily be extended to the full set of rules; some details for disjunction are given in 4.6. So we consider

$$u \mid \lambda v M \mid \exists^+ r M \mid MN \mid M(v.N);$$

in these expressions the variable v gets bound.

We reserve the letters E, F, G for *eliminations*, i.e. expressions of the form $(v.N)$, and R, S, T for both terms and eliminations. Using this notation we obtain a second (and clearly equivalent) inductive definition of terms:

$$u\vec{M} \mid u\vec{M}E \mid \lambda v M \mid \exists^+ r M \mid$$
$$(\lambda v M)N\vec{R} \mid \exists^+ r M(v.N)\vec{R} \mid u\vec{M}ER\vec{S}.$$

derivation	term
$\dfrac{\begin{array}{c}\mid M\\ A\end{array}}{A \vee B}\vee_0^+ \qquad \dfrac{\begin{array}{c}\mid M\\ B\end{array}}{A \vee B}\vee_1^+$	$\left(\vee_{0,B}^+ M^A\right)^{A\vee B} \quad \left(\vee_{1,A}^+ M^B\right)^{A\vee B}$
$\dfrac{\begin{array}{ccc} & [u\colon A] & [v\colon B]\\ \mid M & \mid N & \mid K\\ A \vee B & C & C\end{array}}{C}\vee^- u, v$	$\left(M^{A\vee B}(u^A.N^C, v^B.K^C)\right)^C$
$\dfrac{\begin{array}{cc}\mid M & \mid N\\ A & B\end{array}}{A \wedge B}\wedge^+$	$\langle M^A, N^B\rangle^{A\wedge B}$
$\dfrac{\begin{array}{cc} [u\colon A] & [v\colon B]\\ \mid M & \mid N\\ A \wedge B & C\end{array}}{C}\wedge^- u, v$	$\left(M^{A\wedge B}(u^A, v^B.N^C)\right)^C$
$\dfrac{\begin{array}{cc} & \mid M\\ r & A[x := r]\end{array}}{\exists x A}\exists^+$	$\left(\exists_{x,A}^+ r M^{A[x:=r]}\right)^{\exists x A}$
$\dfrac{\begin{array}{ccc} & [u\colon A] & \\ \mid M & \mid N & \\ \exists x A & B & \end{array}}{B}\exists^- x, u$ (var.cond.)	$\left(M^{\exists x A}(u^A.N^B)\right)^B$ (var.cond.)

TABLE 3. Derivation terms for $\vee$, $\wedge$ and $\exists$

Here the final three forms are not normal: $(\lambda v M)N\vec{R}$ and $\exists^+ r M(v.N)\vec{R}$ both are *β-redexes*, and $u\vec{M}ER\vec{S}$ is a *permutative redex*. The conversion rules are

$$\begin{array}{lll} (\lambda v M)N & \mapsto_\beta M[v := N] & \beta_\to\text{-conversion},\\ \exists_{x,A}^+ r M(v.N) & \mapsto_\beta N[x := r][v := M] & \beta_\exists\text{-conversion},\\ M(v.N)R & \mapsto_\pi M(v.NR) & \text{permutative conversion}. \end{array}$$

The *closure* of these conversions is defined by

- If $M \mapsto_\beta M'$ or $M \mapsto_\pi M'$, then $M \to M'$.
- If $M \to M'$, then also $MR \to M'R$, $NM \to NM'$, $N(v.M) \to N(v.M')$, $\lambda vM \to \lambda vM'$, $\exists^+ rM \to \exists^+ rM'$ (*inner reductions*).

We now give the rules to inductively generate a set SN:

$$\frac{\vec{M} \in \mathsf{SN}}{u\vec{M} \in \mathsf{SN}}\,(\mathsf{Var}_0) \qquad \frac{M \in \mathsf{SN}}{\lambda vM \in \mathsf{SN}}\,(\lambda) \qquad \frac{M \in \mathsf{SN}}{\exists^+ rM \in \mathsf{SN}}\,(\exists)$$

$$\frac{\vec{M}, N \in \mathsf{SN}}{u\vec{M}(v.N) \in \mathsf{SN}}\,(\mathsf{Var}) \qquad \frac{u\vec{M}(v.NR)\vec{S} \in \mathsf{SN}}{u\vec{M}(v.N)R\vec{S} \in \mathsf{SN}}\,(\mathsf{Var}_\pi)$$

$$\frac{M[v := N]\vec{R} \in \mathsf{SN} \qquad N \in \mathsf{SN}}{(\lambda vM)N\vec{R} \in \mathsf{SN}}\,(\beta_\to)$$

$$\frac{N[x := r][v := M]\vec{R} \in \mathsf{SN} \qquad M \in \mathsf{SN}}{\exists^+_{x,A} rM(v.N)\vec{R} \in \mathsf{SN}}\,(\beta_\exists)$$

where in (Var_π) we require that v is not free in R.

Write $M{\downarrow}$ to mean that M is strongly normalizable, i.e., that every reduction sequence starting from M terminates. By analyzing the possible reduction steps we now show that the set $\mathsf{Wf} := \{\, M \mid M{\downarrow} \,\}$ has the closure properties of the definition of SN above, and hence $\mathsf{SN} \subseteq \mathsf{Wf}$.

LEMMA. *Every term in* SN *is strongly normalizable.*

PROOF. We distinguish cases according to the generation rule of SN applied last. The following rules deserve special attention.

Case (Var_π). We prove, as an auxiliary lemma, that

$$u\vec{M}(v.NR)\vec{S}{\downarrow} \text{ implies } u\vec{M}(v.N)R\vec{S}{\downarrow},$$

by induction on $u\vec{M}(v.NR)\vec{S}{\downarrow}$ (i.e., on the reduction tree of this term). We consider the possible reducts of $u\vec{M}(v.N)R\vec{S}$. The only interesting case is $R\vec{S} = (v'.N')T\vec{T}$ and we have a permutative conversion of $R = (v'.N')$ with T, leading to the term $M = u\vec{M}(v.N)(v'.N'T)\vec{T}$. Now $M{\downarrow}$ follows, since

$$u\vec{M}(v.NR)\vec{S} = u\vec{M}(v.N(v'.N'))T\vec{T}$$

leads in two permutative steps to $u\vec{M}(v.N(v'.N'T))\vec{T}$, hence for this term we have the induction hypothesis available.

Case ($\beta_\to$). We show that $M[v := N]\vec{R}{\downarrow}$ and $N{\downarrow}$ imply $(\lambda vM)N\vec{R}{\downarrow}$. This is done by a induction on $N{\downarrow}$, with a side induction on $M[v := N]\vec{R}{\downarrow}$. We need to consider all possible reducts of $(\lambda vM)N\vec{R}$. In case of an outer β-reduction use the assumption. If N is reduced, use the induction hypothesis. Reductions in M and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the side induction hypothesis.

Case ($\beta_\exists$). We show that $N[x := r][v := M]\vec{R}{\downarrow}$ and $M{\downarrow}$ together imply $\exists^+ rM(v.N)\vec{R}{\downarrow}$. This is done by a threefold induction: first on $M{\downarrow}$, second on $N[x := r][v := M]\vec{R}{\downarrow}$ and third on the length of $\vec{R}$. We need to consider all possible reducts of $\exists^+ rM(v.N)\vec{R}$. In case of an outer β-reduction use the

assumption. If M is reduced, use the first induction hypothesis. Reductions in N and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the second induction hypothesis. The only remaining case is when $\vec{R} = S\vec{S}$ and $(v.N)$ is permuted with S, to yield $\exists^+ rM(v.NS)\vec{S}$. Apply the third induction hypothesis, since $(NS)[x := r][v := M]\vec{S} = N[x := r][v := M]S\vec{S}$. □

For later use we prove a slightly generalized form of the rule (Var_π):

Proposition. *If $M(v.NR)\vec{S} \in \mathsf{SN}$, then $M(v.N)R\vec{S} \in \mathsf{SN}$.*

Proof. Induction on the generation of $M(v.NR)\vec{S} \in \mathsf{SN}$. We distinguish cases according to the form of M.

Case $u\vec{T}(v.NR)\vec{S} \in \mathsf{SN}$. If $\vec{T} = \vec{M}$, use (Var_π). Otherwise we have $u\vec{M}(v'.N')\vec{R}(v.NR)\vec{S} \in \mathsf{SN}$. This must be generated by repeated applications of (Var_π) from $u\vec{M}(v'.N'\vec{R}(v.NR)\vec{S}) \in \mathsf{SN}$, and finally by (Var) from $\vec{M} \in \mathsf{SN}$ and $N'\vec{R}(v.NR)\vec{S} \in \mathsf{SN}$. The induction hypothesis for the latter yields $N'\vec{R}(v.N)R\vec{S} \in \mathsf{SN}$, hence $u\vec{M}(v.N'\vec{R}(v.N)R\vec{S}) \in \mathsf{SN}$ by (Var) and finally $u\vec{M}(v.N')\vec{R}(v.N)R\vec{S} \in \mathsf{SN}$ by (Var_π).

Case $\exists^+ rM\vec{T}(v.NR)\vec{S} \in \mathsf{SN}$. Similarly, with $(\beta_\exists)$ instead of (Var_π). In detail: If $\vec{T}$ is empty, by $(\beta_\exists)$ this came from $(NR)[x := r][v := M]\vec{S} = N[x := r][v := M]R\vec{S} \in \mathsf{SN}$ and $M \in \mathsf{SN}$, hence $\exists^+ rM(v.N)R\vec{S} \in \mathsf{SN}$ again by $(\beta_\exists)$. Otherwise we have $\exists^+ rM(v'.N')\vec{T}(v.NR)\vec{S} \in \mathsf{SN}$. This must be generated by $(\beta_\exists)$ from $N'[x := r][v' := M]\vec{T}(v.NR)\vec{S} \in \mathsf{SN}$. The induction hypothesis yields $N'[x := r][v' := M]\vec{T}(v.N)R\vec{S} \in \mathsf{SN}$, hence $\exists^+ rM(v'.N')\vec{T}(v.N)R\vec{S} \in \mathsf{SN}$ by $(\beta_\exists)$.

Case $(\lambda vM)N'\vec{R}(w.NR)\vec{S} \in \mathsf{SN}$. By $(\beta_\to)$ this came from $N' \in \mathsf{SN}$ and $M[v := N']\vec{R}(w.NR)\vec{S} \in \mathsf{SN}$. The induction hypothesis yields $M[v := N']\vec{R}(w.N)R\vec{S} \in \mathsf{SN}$, hence $(\lambda vM)N'\vec{R}(w.N)R\vec{S} \in \mathsf{SN}$ by $(\beta_\to)$. □

In what follows we shall show that *every* term is in SN and hence is strongly normalizable. Given the definition of SN we only have to show that SN is closed under $\to^-$ and $\exists^-$. In order to prove this we must prove simultaneously the closure of SN under substitution.

Theorem (Properties of SN). *For all formulas A,*

(a) *for all $M \in \mathsf{SN}$, if M proves $A = A_0 \to A_1$ and $N \in \mathsf{SN}$, then $MN \in \mathsf{SN}$,*
(b) *for all $M \in \mathsf{SN}$, if M proves $A = \exists xB$ and $N \in \mathsf{SN}$, then $M(v.N) \in \mathsf{SN}$,*
(c) *for all $M \in \mathsf{SN}$, if $N^A \in \mathsf{SN}$, then $M[v := N] \in \mathsf{SN}$.*

Proof. Induction on $\mathsf{dp}(A)$. We prove (a) and (b) before (c), and hence have (a) and (b) available for the proof of (c). More formally, by induction on A we simultaneously prove that (a) holds, that (b) holds and that (a), (b) together imply (c).

(a). By induction on $M \in \mathsf{SN}$. Let $M \in \mathsf{SN}$ and assume that M proves $A = A_0 \to A_1$ and $N \in \mathsf{SN}$. We distinguish cases according to how $M \in \mathsf{SN}$ was generated. For (Var_0), (Var_π), $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again.

Case $u\vec{M}(v.N') \in \mathsf{SN}$ by (Var) from $\vec{M}, N' \in \mathsf{SN}$. Then $N'N \in \mathsf{SN}$ by side induction hypothesis for N', hence $u\vec{M}(v.N'N) \in \mathsf{SN}$ by (Var), hence $u\vec{M}(v.N')N \in \mathsf{SN}$ by (Var_π).

Case $(\lambda v M)^{A_0 \to A_1} \in \mathsf{SN}$ by (λ) from $M \in \mathsf{SN}$. Use $(\beta_\to)$; for this we need to know $M[v := N] \in \mathsf{SN}$. But this follows from IH(c) for M, since N derives A_0.

(b). By induction on $M \in \mathsf{SN}$. Let $M \in \mathsf{SN}$ and assume that M proves $A = \exists x B$ and $N \in \mathsf{SN}$. The goal is $M(v.N) \in \mathsf{SN}$. We distinguish cases according to how $M \in \mathsf{SN}$ was generated. For (Var_π), $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again.

Case $u\vec{M} \in \mathsf{SN}$ by (Var_0) from $\vec{M} \in \mathsf{SN}$. Use (Var).

Case $(\exists^+ r M)^{\exists x A} \in \mathsf{SN}$ by $(\exists)$ from $M \in \mathsf{SN}$. Use $(\beta_\exists)$; for this we need to know $N[x := r][v := M] \in \mathsf{SN}$. But this follows from IH(c) for $N[x := r]$, since M derives $A[x := r]$.

Case $u\vec{M}(v'.N') \in \mathsf{SN}$ by (Var) from $\vec{M}, N' \in \mathsf{SN}$. Then $N'(v.N) \in \mathsf{SN}$ by side induction hypothesis for N', hence $u\vec{M}(v.N'(v.N)) \in \mathsf{SN}$ by (Var) and therefore $u\vec{M}(v.N')(v.N) \in \mathsf{SN}$ by (Var_π).

(c). By induction on $M \in \mathsf{SN}$. Let $N^A \in \mathsf{SN}$; the goal is $M[v := N] \in \mathsf{SN}$. We distinguish cases according to how $M \in \mathsf{SN}$ was generated. For (λ), $(\exists)$, $(\beta_\to)$ and $(\beta_\exists)$ use the same rule again.

Case $u\vec{M} \in \mathsf{SN}$ by (Var_0) from $\vec{M} \in \mathsf{SN}$. Then $\vec{M}[v := N] \in \mathsf{SN}$ by SIH(c). If $u \neq v$, use (Var_0) again. If $u = v$, we must show $N\vec{M}[v := N] \in \mathsf{SN}$. Note that N proves A; hence the claim follows from (a) and the induction hypothesis.

Case $u\vec{M}(v'.N') \in \mathsf{SN}$ by (Var) from $\vec{M}, N' \in \mathsf{SN}$. If $u \neq v$, use (Var) again. If $u = v$, we must show $N\vec{M}[v := N](v'.N'[v := N]) \in \mathsf{SN}$. Note that N proves A; hence in case $\vec{M}$ empty the claim follows from (b), and otherwise from (a) and the induction hypothesis.

Case $u\vec{M}(v'.N')R\vec{S} \in \mathsf{SN}$ by (Var_π) from $u\vec{M}(v'.N'R)\vec{S} \in \mathsf{SN}$. If $u \neq v$, use (Var_π) again. If $u = v$, from the induction hypothesis we obtain

$$N\vec{M}[v := N](v'.N'[v := N]R[v := N]).\vec{S}[v := N] \in \mathsf{SN}$$

Now use the proposition above. □

Corollary. *Every term is strongly normalizable.*

Proof. Induction on the (first) inductive definition of terms M. In cases u and $\lambda v M$ the claim follows from the definition of SN, and in cases MN and $M(v.N)$ it follows from parts (a), (b) of the previous theorem. □

4.6. Disjunction. We describe the changes necessary to extend the result above to the language with disjunction $\vee$.

We have additional β-conversions

$$\vee_i^+ M(v_0.N_0, v_1.N_1) \mapsto_\beta M[v_i := N_i] \quad \beta_{\vee_i}\text{-conversion.}$$

The definition of SN needs to be extended by

$$\frac{M \in \mathsf{SN}}{\vee_i^+ M \in \mathsf{SN}}\,(\vee_i)$$

$$\frac{\vec{M}, N_0, N_1 \in \mathsf{SN}}{u\vec{M}(v_0.N_0, v_1.N_1) \in \mathsf{SN}}\,(\mathsf{Var}_\vee) \quad \frac{u\vec{M}(v_0.N_0R, v_1.N_1R)\vec{S} \in \mathsf{SN}}{u\vec{M}(v_0.N_0, v_1.N_1)R\vec{S} \in \mathsf{SN}}\,(\mathsf{Var}_{\vee,\pi})$$

$$\frac{N_i[v_i := M]\vec{R} \in \mathsf{SN} \qquad N_{1-i}\vec{R} \in \mathsf{SN} \qquad M \in \mathsf{SN}}{\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R} \in \mathsf{SN}}\,(\beta_{\vee_i})$$

The former rules (Var), (Var_π) should then be renamed into ($\mathsf{Var}_\exists$), ($\mathsf{Var}_{\exists,\pi}$).

The lemma above stating that every term in SN is strongly normalizable needs to be extended by an additional clause:

Case ($\beta_{\vee_i}$). We show that $N_i[v_i := M]\vec{R}\downarrow$, $N_{1-i}\vec{R}\downarrow$ and $M\downarrow$ together imply $\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R}\downarrow$. This is done by a fourfold induction: first on $M\downarrow$, second on $N_i[v_i := M]\vec{R}\downarrow$, $N_{1-i}\vec{R}\downarrow$, third on $N_{1-i}\vec{R}\downarrow$ and fourth on the length of $\vec{R}$. We need to consider all possible reducts of $\vee_i^+ M(v_0.N_0, v_1.N_1)\vec{R}$. In case of an outer β-reduction use the assumption. If M is reduced, use the first induction hypothesis. Reductions in N_i and in $\vec{R}$ as well as permutative reductions within $\vec{R}$ are taken care of by the second induction hypothesis. Reductions in N_{1-i} are taken care of by the third induction hypothesis. The only remaining case is when $\vec{R} = S\vec{S}$ and $(v_0.N_0, v_1.N_1)$ is permuted with S, to yield $(v_0.N_0S, v_1.N_1S)$. Apply the fourth induction hypothesis, since $(N_iS)[v := M]\vec{S} = N_i[v := M]S\vec{S}$.

Finally the theorem above stating properties of SN needs an additional clause:

- for all $M \in \mathsf{SN}$, if M proves $A = A_0 \vee A_1$ and $N_0, N_1 \in \mathsf{SN}$, then $M(v_0.N_0, v_1.N_1) \in \mathsf{SN}$.

Proof. The new clause is proved by induction on $M \in \mathsf{SN}$. Let $M \in \mathsf{SN}$ and assume that M proves $A = A_0 \vee A_1$ and $N_0, N_1 \in \mathsf{SN}$. The goal is $M(v_0.N_0, v_1.N_1) \in \mathsf{SN}$. We distinguish cases according to how $M \in \mathsf{SN}$ was generated. For ($\mathsf{Var}_{\exists,\pi}$), ($\mathsf{Var}_{\vee,\pi}$), ($\beta_\to$), ($\beta_\exists$) and ($\beta_{\vee_i}$) use the same rule again.

Case $u\vec{M} \in \mathsf{SN}$ by (Var_0) from $\vec{M} \in \mathsf{SN}$. Use ($\mathsf{Var}_\vee$).

Case $(\vee_i^+ M)^{A_0 \vee A_1} \in \mathsf{SN}$ by ($\vee_i$) from $M \in \mathsf{SN}$. Use ($\beta_{\vee_i}$); for this we need to know $N_i[v_i := M] \in \mathsf{SN}$ and $N_{1-i} \in \mathsf{SN}$. The latter is assumed, and the former follows from main induction hypothesis (with N_i) for the substitution clause of the theorem, since M derives A_i.

Case $u\vec{M}(v'.N') \in \mathsf{SN}$ by ($\mathsf{Var}_\exists$) from $\vec{M}, N' \in \mathsf{SN}$. For brevity let $E := (v_0.N_0, v_1.N_1)$. Then $N'E \in \mathsf{SN}$ by side induction hypothesis for N', so $u\vec{M}(v'.N'E) \in \mathsf{SN}$ by ($\mathsf{Var}_\exists$) and therefore $u\vec{M}(v'.N')E \in \mathsf{SN}$ by ($\mathsf{Var}_{\exists,\pi}$).

Case $u\vec{M}(v_0'.N_0', v_1'.N_1') \in \mathsf{SN}$ by ($\mathsf{Var}_\vee$) from $\vec{M}, N_0', N_1' \in \mathsf{SN}$. Let $E := (v_0.N_0, v_1.N_1)$. Then $N_i'E \in \mathsf{SN}$ by side induction hypothesis for N_i', so $u\vec{M}(v_0'.N_0'E, v_1'.N_1'E) \in \mathsf{SN}$ by ($\mathsf{Var}_\vee$) and therefore $u\vec{M}(v_0'.N_0', v_1'.N_1')E \in \mathsf{SN}$ by ($\mathsf{Var}_{\vee,\pi}$).

Clause (c) now needs additional cases, e.g.,

Case $u\vec{M}(v_0.N_0, v_1.N_1) \in \mathsf{SN}$ by ($\mathsf{Var}_\vee$) from $\vec{M}, N_0, N_1 \in \mathsf{SN}$. If $u \neq v$, use ($\mathsf{Var}_\vee$). If $u = v$, we show $N\vec{M}[v := N](v_0.N_0[v := N], v_1.N_1[v := N]) \in \mathsf{SN}$. Note that N proves A; hence in case $\vec{M}$ empty the claim follows from (b), and otherwise from (a) and the induction hypothesis. □

4.7. The Structure of Normal Derivations. As mentioned already, normalizations aim at removing local maxima of complexity, i.e. formula occurrences which are first introduced and immediately afterwards eliminated.

However, an introduced formula may be used as a minor premise of an application of $\vee^-$, $\wedge^-$ or $\exists^-$, then stay the same throughout a sequence of applications of these rules, being eliminated at the end. This also constitutes a local maximum, which we should like to eliminate; for that we need the so-called permutative conversions. First we give a precise definition.

DEFINITION. A *segment* of (*length* n) in a derivation M is a sequence $A_1, \ldots, A_n$ of occurrences of a formula A such that

(a) for $1 < i < n$, A_i is a minor premise of an application of $\vee^-$, $\wedge^-$ or $\exists^-$, with conclusion A_{i+1};
(b) A_n is not a minor premise of $\vee^-$, $\wedge^-$ or $\exists^-$.
(c) A_1 is not the conclusion of $\vee^-$, $\wedge^-$ or $\exists^-$.

(Note: An f.o. which is neither a minor premise nor the conclusion of an application of $\vee^-$, $\wedge^-$ or $\exists^-$ always belongs to a segment of length 1.) A segment is *maximal* or a *cut (segment)* if A_n is the major premise of an E-rule, and either $n > 1$, or $n = 1$ and $A_1 = A_n$ is the conclusion of an I-rule.

We shall use σ, σ' for segments. We shall say that σ is a *subformula* of σ' if the formula A in σ is a subformula of B in σ'. Clearly a derivation is normal iff it does not contain a maximal segment.

The argument in 3.7 needs to be refined to also cover the rules for $\vee, \wedge, \exists$. The reason for the difficulty is that in the E-rules $\vee^-, \wedge^-, \exists^-$ the subformulas of a major premise $A \vee B$, $A \wedge B$ or $\exists x A$ of an E-rule application do not appear in the conclusion, but among the assumptions being discharged by the application. This suggests the definition of track below.

The general notion of a track is designed to retain the subformula property in case one passes through the major premise of an application of a $\vee^-, \wedge^-, \exists^-$-rule. In a track, when arriving at an A_i which is the major premise of an application of such a rule, we take for A_{i+1} a hypothesis discharged by this rule.

DEFINITION. A *track* of a derivation M is a sequence of f.o.'s $A_0, \ldots, A_n$ such that

(a) A_0 is a top f.o. in M not discharged by an application of an $\vee^-, \wedge^-, \exists^-$-rule;
(b) A_i for $i < n$ is not the minor premise of an instance of $\rightarrow^-$, and *either*
  (i) A_i is not the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule and A_{i+1} is directly below A_i, *or*
  (ii) A_i is the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule and A_{i+1} is an assumption discharged by this instance;
(c) A_n is *either*
  (i) the minor premise of an instance of $\rightarrow^-$, *or*
  (ii) the conclusion of M, *or*
  (iii) the major premise of an instance of a $\vee^-, \wedge^-, \exists^-$-rule in case there are no assumptions discharged by this instance.

PROPOSITION. *Let M be a normal derivation, and let $\pi = \sigma_0, \ldots, \sigma_n$ be a track in M. Then there is a segment σ_i in π, the* minimum *segment or* minimum *part of the track, which separates two (possibly empty) parts of π,*

*called the E-part (*elimination part*) and the I-part (*introduction part*) of* π *such that*

(a) *for each* σ_j *in the E-part one has* $j < i$, σ_j *is a major premise of an E-rule, and* σ_{j+1} *is a strictly positive part of* σ_j, *and therefore each* σ_j *is a s.p.p. of* σ_0*;*

(b) *for each* σ_j *which is the minimum segment or is in the I-part one has* $i \leq j$, *and if* $j \neq n$, *then* σ_j *is a premise of an I-rule and a s.p.p. of* σ_{j+1}, *so each* σ_j *is a s.p.p. of* σ_n.

DEFINITION. A *track of order* 0, or *main track*, in a normal derivation is a track ending either in the conclusion of the whole derivation or in the major premise of an application of a $\vee^-$, $\wedge^-$ or $\exists^-$-rule, provided there are no assumption variables discharged by the application. A *track of order* $n+1$ is a track ending in the minor premise of an $\rightarrow^-$-application, with major premise belonging to a track of order n.

A *main branch* of a derivation is a branch π in the prooftree such that π passes only through premises of I-rules and *major premises* of E-rules, and π begins at a top node and ends in the conclusion.

REMARK. By an obvious *simplification conversion* we may remove every application of an $\vee^-$, $\wedge^-$ or $\exists^-$-rule that discharges no assumption variables. If such simplification conversion are performed, each track of order 0 in a normal derivation is a track ending in the conclusion of the whole derivation.

If we search for a main branch going upwards from the conclusion, the branch to be followed is unique as long as we do not encounter an $\wedge^+$-application.

LEMMA. *In a normal derivation each formula occurrence belongs to some track.*

PROOF. By induction on the height of normal derivations. For example, suppose a derivation K ends with an $\exists^-$-application:

$$\dfrac{\exists x A \qquad \begin{matrix}[u\colon A] \\ \mid N \\ B\end{matrix}}{B}\ \exists^- x, u \qquad \text{with } \mid M \text{ above } \exists x A$$

B in N belongs to a track π (induction hypothesis); either this does not start in $u\colon A$, and then π, B is a track in K which ends in the conclusion; or π starts in $u\colon A$, and then there is a track π' in M (induction hypothesis) such that π', π, C is a track in K ending in the conclusion. The other cases are left to the reader. □

THEOREM (Subformula property). *Let* M *be a normal derivation where every application of an* $\vee^-$, $\wedge^-$ *or* $\exists^-$*-rule discharges at least one assumption variable. Then each formula occurring in the derivation is a subformula of either the end formula or else an assumption formula.*

PROOF. As note above, each track of order 0 in M is a track ending in the conclusion of M. We can now prove the theorem for tracks of order n, by induction on n. □

THEOREM (Disjunction property). *If* Γ *does not contain a disjunction as s.p.p. (= strictly positive part, defined in 1.3), then, if* $\Gamma \vdash A \vee B$*, it follows that* $\Gamma \vdash A$ *or* $\Gamma \vdash B$*.*

PROOF. Consider a normal derivation M of $A \vee B$ from assumptions Γ not containing a disjunction as s.p.p. The conclusion $A \vee B$ is the final formula of a (main) track, whose top formula A_0 in M must be an assumption in Γ. Since Γ does not contain a disjunction as s.p.p., the segment σ with the conclusion $A \vee B$ is in the I-part. Skip the final $\vee_i^+$-rule and replace the formulas in σ by A if $i = 0$, and by B if $i = 1$. □

There is a similar theorem for the existential quantifier:

THEOREM (Explicit definability under hypotheses). *Let* $\Gamma \vdash \exists x A$*.*

(a) *If* Γ *does not contain an existential s.p.p., then there are terms* r_1*,* r_2*, . . . ,* r_n *such that* $\Gamma \vdash A[x := r_1] \vee \ldots \vee A[x := r_n]$*.*

(b) *If* Γ *neither contains a disjunctive s.p.p., nor an existential s.p.p., then there is a term* r *such that* $\Gamma \vdash A[x := r]$*.*

PROOF. Consider a normal derivation M of $\exists x A$ from assumptions Γ not containing an existential s.p.p. We use induction on the derivation, and distinguish cases on the last rule.

(a). By assumption the last rule cannot be $\exists^-$. We only consider the case $\vee^-$ and leave the others to the reader.

$$\dfrac{\begin{array}{ccc} & [u\colon B] & [v\colon C] \\ \mid M & \mid N_0 & \mid N_1 \\ B \vee C & \exists x A & \exists x A \end{array}}{\exists x A}\ \vee^- u, v$$

By assumption again neither B nor C can have an existential s.p.p. Applying the induction hypothesis to N_0 and N_1 we obtain

$$\dfrac{\begin{array}{ccc} & [u\colon B] & [v\colon C] \\ & \mid N_0 & \mid N_1 \\ \mid M & \dfrac{\bigvee_{i=1}^{n} A[x := r_i]}{\bigvee_{i=1}^{n+m} A[x := r_i]}\ \vee^+ & \dfrac{\bigvee_{i=n+1}^{n+m} A[x := r_i]}{\bigvee_{i=1}^{n+m} A[x := r_i]}\ \vee^+ \\ B \vee C & & \end{array}}{\bigvee_{i=1}^{n+m} A[x := r_i]}\ \vee^- u, v$$

(b). Similarly; by assumption the last rule can be neither $\vee^-$ nor $\exists^-$. □

REMARK. *Rasiowa-Harrop formulas* (in the literature also called *Harrop* formulas) are formulas for which no s.p.p. is a disjunction or an existential formula. For Γ consisting of Rasiowa-Harrop formulas both theorems above hold.

5. Notes

The proof of the existence of normal forms w.r.t permutative conversions is originally due to Prawitz [**20**]. We have adapted a method developed by Joachimski and Matthes [**13**], which in turn is based on van Raamsdonk's and Severi's [**28**].

CHAPTER 2

Models

It is an obvious question to ask whether the logical rules we have been considering suffice, i.e. whether we have forgotten some necessary rules. To answer this question we first have to fix the *meaning* of a formula, i.e. we have to provide a semantics.

This is rather straightforward for classical logic: we can take the usual notion of a structure (or model, or (universal) algebra). However, for minimal and intuitionistic logic we need a more refined notion: we shall use so-called Beth-structures here. Using this concept of a model we will prove soundness and completeness for both, minimal and intuitionistic logic. As a corollary we will obtain completeness of classical logic, w.r.t. the standard notion of a structure.

1. Structures for Classical Logic

1.1. Structures. We define the notion of a structure (more accurately $\mathcal{L}$-structure) and define what the value of a term and the meaning of a formula in such a structure should be.

DEFINITION. $\mathcal{M} = (D, I)$ is a *pre-structure* (or $\mathcal{L}$-pre-structure), if D a non-empty set (the *carrier set* or the *domain* of $\mathcal{M}$) and I is a map (*interpretation*) assigning to every n-ary function symbol f of $\mathcal{L}$ a function

$$I(f)\colon D^n \to D.$$

In case $n = 0$, $I(f)$ is an element of D. $\mathcal{M} = (D, I_0, I_1)$ is a *structure* (or $\mathcal{L}$-structure), if (D, I_0) is a pre-structure and I_1 a map assigning to every n-ary relation symbol R of $\mathcal{L}$ an n-ary relation

$$I_1(R) \subseteq D^n.$$

In case $n = 0$, $I_1(R)$ is one of the truth values 1 and 0; in particular we require $I_1(\bot) = 0$.

If $\mathcal{M} = (D, I)$ or (D, I_0, I_1), then we often write $|\mathcal{M}|$ for the carrier set D of $\mathcal{M}$ and $f^{\mathcal{M}}$, $R^{\mathcal{M}}$ for the interpretations $I_0(f)$, $I_1(R)$ of the function and relation symbols.

An *assignment* (or variable assignment) in D is a map η assigning to every variable $x \in \mathsf{dom}(\eta)$ a value $\eta(x) \in D$. Finite assignments will be written as $[x_1 := a_1, \ldots, x_n := a_n]$ (or else as $[a_1/x_1, \ldots, a_n/x_n]$), with distinct $x_1, \ldots, x_n$. If η is an assignment in D and $a \in D$, let η^a_x be the assignment in D mapping x to a and coinciding with η elsewhere, so

$$\eta^a_x(y) := \begin{cases} \eta(y), & \text{if } y \neq x \\ a, & \text{if } y = x. \end{cases}$$

Let a pre-structure $\mathcal{M}$ and an assignment η in $|\mathcal{M}|$ be given. We define a homomorphic extension of η (denoted by η as well) to the set Set $\mathsf{Ter}_{\mathcal{L}}$ of $\mathcal{L}$-terms t such that $\mathsf{vars}(t) \subseteq \mathsf{dom}(\eta)$ by

$$\begin{aligned} \eta(c) &:= c^{\mathcal{M}}, \\ \eta(f(t_1, \ldots, t_n)) &:= f^{\mathcal{M}}(\eta(t_1), \ldots, \eta(t_n)). \end{aligned}$$

Observe that the extension of η depends on $\mathcal{M}$; therefore we may also write $t^{\mathcal{M}}[\eta]$ for $\eta(t)$.

For every structure $\mathcal{M}$, assignment η in $|\mathcal{M}|$ and formula A with $\mathsf{FV}(A) \subseteq \mathsf{dom}(\eta)$ we define $\mathcal{M} \models A[\eta]$ (read: A is *valid* in $\mathcal{M}$ under the assignment η) by recursion on A.

$$\begin{aligned} \mathcal{M} \models R(t_1, \ldots, t_n)[\eta] &:\Longleftrightarrow (t_1^{\mathcal{M}}[\eta], \ldots, t_n^{\mathcal{M}}[\eta]) \in I_1(R) \quad \text{for } R \text{ not 0-ary.} \\ \mathcal{M} \models R[\eta] &:\Longleftrightarrow I_1(R) = 1 \quad \text{for } R \text{ 0-ary.} \\ \mathcal{M} \models (A \wedge B)[\eta] &:\Longleftrightarrow \mathcal{M} \models A[\eta] \text{ and } \mathcal{M} \models B[\eta]. \\ \mathcal{M} \models (A \vee B)[\eta] &:\Longleftrightarrow \mathcal{M} \models A[\eta] \text{ or } \mathcal{M} \models B[\eta]. \\ \mathcal{M} \models (A \to B)[\eta] &:\Longleftrightarrow \text{if } \mathcal{M} \models A[\eta], \text{ then } \mathcal{M} \models B[\eta]. \\ \mathcal{M} \models (\forall x A)[\eta] &:\Longleftrightarrow \text{for all } a \in |\mathcal{M}| \text{ we have } \mathcal{M} \models A[\eta_x^a]. \\ \mathcal{M} \models (\exists x A)[\eta] &:\Longleftrightarrow \text{there is an } a \in |\mathcal{M}| \text{ such that } \mathcal{M} \models A[\eta_x^a]. \end{aligned}$$

Because of $I_1(\bot) = 0$ we have in particular $\mathcal{M} \not\models \bot[\eta]$.

If Γ is a set of formulas, we write $\mathcal{M} \models \Gamma[\eta]$, if for all $A \in \Gamma$ we have $\mathcal{M} \models A[\eta]$. If $\mathcal{M} \models A[\eta]$ for all assignments η in $|\mathcal{M}|$, we write $\mathcal{M} \models A$.

1.2. Coincidence and Substitution Lemma.

Lemma (Coincidence). *Let $\mathcal{M}$ be a structure, t a term, A a formula and η, ξ assignments in $|\mathcal{M}|$.*

(a) *If $\eta(x) = \xi(x)$ for all $x \in \mathsf{vars}(t)$, then $\eta(t) = \xi(t)$.*
(b) *If $\eta(x) = \xi(x)$ for all $x \in \mathsf{FV}(A)$, then $\mathcal{M} \models A[\eta]$ iff $\mathcal{M} \models A[\xi]$.*

Proof. Induction on terms and formulas. □

Lemma (Substitution). *Let $\mathcal{M}$ be an $\mathcal{L}$-structure, t, r $\mathcal{L}$-terms, A an $\mathcal{L}$-formula and η an assignment in $|\mathcal{M}|$. Then*

(a) $\eta(r[x := t]) = \eta_x^{\eta(t)}(r)$.
(b) $\mathcal{M} \models A[x := t][\eta] \iff \mathcal{M} \models A[\eta_x^{\eta(t)}]$.

Proof. (a). Induction on r. (b). Induction on A. We restrict ourselves to the cases of an atomic formula and a universal formula; the other cases are easier.

Case $R(s_1, \ldots, s_n)$. For simplicity assume $n = 1$. Then

$$\begin{aligned} \mathcal{M} \models R(s)[x := t][\eta] &\iff \mathcal{M} \models R(s[x := t])[\eta] \\ &\iff \eta(s[x := t]) \in R^{\mathcal{M}} \\ &\iff \eta_x^{\eta(t)}(s) \in R^{\mathcal{M}} \quad \text{by (a)} \\ &\iff \mathcal{M} \models R(s)[\eta_x^{\eta(t)}]. \end{aligned}$$

Case $\forall y A$. We may assume $y \neq x$ and $y \notin \mathsf{vars}(t)$.

$$\begin{aligned}
&\mathcal{M} \models (\forall y A)[x := t][\eta] \\
&\iff \mathcal{M} \models (\forall y A[x := t])[\eta] \\
&\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M} \models A[x := t][\eta^a_y] \\
&\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M} \models A[(\eta^a_y)^b_x] \text{ with } b := \eta^a_y(t) = \eta(t) \\
&\quad \text{(by IH and the coincidence lemma)} \\
&\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M} \models A[(\eta^b_x)^a_y], \quad \text{(because } x \neq y\text{)} \\
&\iff \mathcal{M} \models (\forall y A)[\eta^b_x]
\end{aligned}$$

This completes the proof. □

1.3. Soundness. We prove the soundness theorem: it says that every formula derivable in classical logic is valid in an arbitrary structure.

THEOREM (Soundness). *Let* $\Gamma \vdash_c B$. *If* $\mathcal{M}$ *is a structure and* η *an assignment in* $|\mathcal{M}|$, *then* $\mathcal{M} \models \Gamma[\eta]$ *entails* $\mathcal{M} \models B[\eta]$.

PROOF. Induction on derivations. The given derivation of B from Γ can only have finitely many free assumptions; hence we may assume $\Gamma = \{A_1, \dots, A_n\}$.

Case $u\colon B$. Then $B \in \Gamma$ and the claim is obvious.

Case $\mathsf{Stab}_R\colon \forall \vec{x}.\neg\neg R\vec{x} \to R\vec{x}$. Again the claim is clear, since $\mathcal{M} \models \neg\neg A[\eta]$ means the same as $\mathcal{M} \models A[\eta]$.

Case $\to^-$. Assume $\mathcal{M} \models \Gamma[\eta]$. We must show $\mathcal{M} \models B[\eta]$. By IH, $\mathcal{M} \models (A \to B)[\eta]$ and $\mathcal{M} \models A[\eta]$. The claim follows from the definition of $\models$.

Case $\to^+$. Assume $\mathcal{M} \models \Gamma[\eta]$. We must show $\mathcal{M} \models (A \to B)[\eta]$. So assume in addition $\mathcal{M} \models A[\eta]$. We must show $\mathcal{M} \models B[\eta]$. By IH (with $\Gamma \cup \{A\}$ instead of Γ) this clearly holds.

Case $\forall^+$. Assume $\mathcal{M} \models \Gamma[\eta]$. We must show $\mathcal{M} \models A[\eta^a_x]$. We may assume that all assumptions $A_1, \dots, A_n$ actually in the given derivation. Since because of the variable condition for $\forall^+$ the variable x does not appear free in any of the formulas $A_1, \dots, A_n$, we have by the coincidence lemma $\mathcal{M} \models \Gamma[\eta^a_x]$. The IH (with η^a_x instead of η) yields $\mathcal{M} \models A[\eta^a_x]$.

Case $\forall^-$. Assume $\mathcal{M} \models \Gamma[\eta]$. We must show $\mathcal{M} \models A[x := t][\eta]$, i.e. by the substitution lemma $\mathcal{M} \models A[\eta^b_x]$ with $b := \eta(t)$. By IH, $\mathcal{M} \models (\forall x A)[\eta]$, i.e. $\mathcal{M} \models A[\eta^a_x]$ for all $a \in |\mathcal{M}|$. With $\eta(t)$ for a the claim follows.

The other cases are proved similarly. □

2. Beth-Structures for Minimal Logic

2.1. Beth-Structures. Consider a partially ordered set of "possible worlds". The worlds are represented as nodes in a finitely branching tree. They may be thought of as possible states such that all nodes "above" a node k are the ways in which k may develop in the future. The worlds are increasing, that is, if an atomic formula $R\vec{t}$ true is in a world k, then $R\vec{t}$ is true in all worlds "above" k.

More formally, each Beth-structure is based on a finitely branching tree T. A node k over a set S is a finite sequence $k = \langle a_0, a_1, \dots, a_{n-1}\rangle$ of

elements of S; $\mathsf{lh}(k)$ is the length of k. We write $k \preceq k'$ if k is the initial segment of k'. A tree on S is a set of nodes closed under initial segments. A tree T is finitely branching if every node in T has finitely many immediate successors.

A tree T is *unbounded* if for every $n \in \mathbb{N}$ there is a node $k \in T$ such that $\mathsf{lh}(k) = n$. A *branch* of T is a linearly ordered subtree of T. A *leaf* is a node without successors in T.

For the proof of the completeness theorem, a Beth-structure based on a complete binary tree (i.e. the complete tree over $\{0, 1\}$) will suffice. The nodes will be all the finite sequences of 0's and 1's, and the ordering is as above. The root is the empty sequence and $k0$ is the sequence k with the postfix 0. Similarly for $k1$.

Definition. Let $(T, \preceq)$ be a finitely branching tree. $\mathcal{B} = (D, I_0, I_1)$ is a *$\mathcal{L}$-Beth-structure* on T, where D is a nonempty set, and for each n-ary function symbol in $\mathcal{L}$, I_0 assigns f a map $I_0(f)\colon D^n \to D$. For each n-ary relation symbol R in $\mathcal{L}$ and each node $k \in T$, $I_1(R, k) \subseteq D^n$ is assigned in such a way that monotonicity is preserved, that is,

$$k \preceq k' \Rightarrow I_1(R, k) \subseteq I_1(R, k').$$

If $n = 0$, then $I_1(R, k)$ is either true or false, and it follows by the monotonicity that if $k \preceq k'$ and $I_1(R, k)$ then $I_1(R, k')$.

There is no special requirement set on $I_1(\bot, k)$. In minimal logic, *falsum* $\bot$ plays a role of an ordinary propositional variable.

For an assignment η, $t^{\mathcal{B}}[\eta]$ is understood classically. The classical satisfaction relation $\mathcal{M} \models A[\eta]$ is replaced with the forcing relation in Beth-structures. It is obvious from the definition that any T can be extended to a complete tree $\bar{T}$ without leaves, in which for each leaf $k \in T$ all sequences $k0$, $k00$, $k000, \dots$ are added to T. For each node $k0 \dots 0$, we add $I_1(R, k0 \dots 0) := I_1(R, k)$.

Definition. $\mathcal{B}, k \Vdash A[\eta]$ ($\mathcal{B}$ forces A at a node k for an assignment η) is defined inductively as follows. We write $k \Vdash A[\eta]$ when it is clear from the context what the underlying structure $\mathcal{B}$ is, and we write $\forall k' \succeq_n k\, A$ for $\forall k' \succeq k.\mathsf{lh}(k') = \mathsf{lh}(k) + n \to A$.

$$\begin{aligned}
k \Vdash R(t_1, \dots, t_p)[\eta] \quad &:\Longleftrightarrow \quad \exists n \forall k' \succeq_n k\, (t_1^{\mathcal{B}}[\eta], \dots, t_p^{\mathcal{B}}[\eta]) \in I_1(R, k'), \\
&\qquad \text{if } R \text{ is not 0-ary.} \\
k \Vdash R[\eta] \quad &:\Longleftrightarrow \quad \exists n \forall k' \succeq_n k\, I_1(R, k') = 1 \quad \text{if } R \text{ is 0-ary.} \\
k \Vdash (A \vee B)[\eta] \quad &:\Longleftrightarrow \quad \exists n \forall k' \succeq_n k.k' \Vdash A[\eta] \ \text{ or } \ k' \Vdash B[\eta]. \\
k \Vdash (\exists x A)[\eta] \quad &:\Longleftrightarrow \quad \exists n \forall k' \succeq_n k \exists a \in |\mathcal{B}|\, k' \Vdash A[\eta_x^a]. \\
k \Vdash (A \to B)[\eta] \quad &:\Longleftrightarrow \quad \forall k' \succeq k.k' \Vdash A[\eta] \Rightarrow k' \Vdash B[\eta]. \\
k \Vdash (A \wedge B)[\eta] \quad &:\Longleftrightarrow \quad k \Vdash A[\eta] \ \text{ and } \ k \Vdash B[\eta]. \\
k \Vdash (\forall x A)[\eta] \quad &:\Longleftrightarrow \quad \forall a \in |\mathcal{B}|\, k \Vdash A[\eta_x^a].
\end{aligned}$$

The clauses for atoms, disjunction and existential quantifier include a concept of a "bar", in $\bar{T}$.

2.2. Covering Lemma. It is easily seen (from the definition and using monotonicity) that from $k \Vdash A[\eta]$ and $k \preceq k'$ we can conclude $k' \Vdash A[\eta]$. The converse is also true:

LEMMA (Covering Lemma).

$$\forall k' \succeq_n k \, k' \Vdash A[\eta] \Rightarrow k \Vdash A[\eta].$$

PROOF. Induction on A. We write $k \Vdash A$ for $k \Vdash A[\eta]$.

Case $R\vec{t}$. Assume

$$\exists n \forall k' \succeq_n k \, k' \Vdash R\vec{t},$$

hence by definition

$$\exists n \forall k' \succeq_n k \exists m \forall k'' \succeq_m k' \vec{t}^{\mathcal{B}}[\eta] \in I_1(R, k'').$$

Since T is a finitely branching tree,

$$\exists m \forall k' \succeq_m k \, \vec{t}^{\mathcal{B}}[\eta] \in I_1(R, k').$$

Hence $k \Vdash R\vec{t}$.

The cases $A \vee B$ and $\exists x A$ are handled similarly.

Case $A \to B$. Let $k' \Vdash A \to B$ for all $k' \succeq k$ with $\mathsf{lh}(k') = \mathsf{lh}(k) + n$. We show

$$\forall l \succeq k . l \Vdash A \Rightarrow l \Vdash B.$$

Let $l \succeq k$ and $l \Vdash A$. We show that $l \Vdash B$. We apply the IH to B and $m := \max(\mathsf{lh}(k) + n, \mathsf{lh}(l))$. So assume $l' \succeq l$ and $\mathsf{lh}(l') = m$. It is sufficient to show $l' \Vdash B$. If $\mathsf{lh}(l') = \mathsf{lh}(l)$, then $l' = l$ and we are done. If $\mathsf{lh}(l') = \mathsf{lh}(k) + n > \mathsf{lh}(l)$, then l' is an extension of l as well as of k and length $\mathsf{lh}(k) + n$, and hence $l' \Vdash A \to B$ by assumption. Moreover, $l' \Vdash A$, since $l' \succeq l$ and $l \Vdash A$. It follows that $l' \Vdash B$.

The cases $A \wedge B$ and $\forall x A$ are obvious. □

2.3. Coincidence and Substitution. The coincidence and substitution lemmas hold for Beth-structures.

LEMMA (Coincidence). *Let $\mathcal{B}$ be a Beth-structure, t a term, A a formula and η, ξ assignments in $|\mathcal{B}|$.*

(a) *If $\eta(x) = \xi(x)$ for all $x \in \mathsf{vars}(t)$, then $\eta(t) = \xi(t)$.*
(b) *If $\eta(x) = \xi(x)$ for all $x \in \mathsf{FV}(A)$, then $\mathcal{B}, k \Vdash A[\eta] \iff \mathcal{B}, k \Vdash A[\xi]$.*

PROOF. Induction on terms and formulas. □

LEMMA (Substitution). *Let $\mathcal{B}$ be a Beth-structure, t, r terms, A a formula and η an assignment in $|\mathcal{B}|$. Then*

(a) $\eta(r[x := t]) = \eta_x^{\eta(t)}(r)$.
(b) $\mathcal{B}, k \Vdash A[x := t][\eta] \iff \mathcal{B}, k \Vdash A[\eta_x^{\eta(t)}]$.

PROOF. Induction on terms and formulas. □

2.4. Soundness. As usual, we proceed to prove soundness theorem.

THEOREM (Soundness). *Let $\Gamma \cup \{A\}$ be a set of formulas such that $\Gamma \vdash A$. Then, if $\mathcal{B}$ is a Beth-structure, k a node and η an assignment in $|\mathcal{B}|$, it follows that $\mathcal{B}, k \Vdash \Gamma[\eta]$ entails $\mathcal{B}, k \Vdash A[\eta]$.*

PROOF. Induction on derivations.

We begin with the axiom schemes $\vee_0^+$, $\vee_1^+$, $\vee^-$, $\exists^+$ and $\exists^-$. $k \Vdash C[\eta]$ is abbreviated $k \Vdash C$, when η is known from the context.

Case $\vee_0^+ \colon A \to A \vee B$. We show $k \Vdash A \to A \vee B$. Assume for $k' \succeq k$ that $k' \Vdash A$. Show: $k' \Vdash A \vee B$. This follows from the definition, since $k' \Vdash A$. The case $\vee_1^+ \colon B \to A \vee B$ is symmetric.

Case $\vee^- \colon A \vee B \to (A \to C) \to (B \to C) \to C$. We show that $k \Vdash A \vee B \to (A \to C) \to (B \to C) \to C$. Assume for $k' \succeq k$ that $k' \Vdash A \vee B$, $k' \Vdash A \to C$ and $k' \Vdash B \to C$ (we can safely assume that k' is the same for all three premises). Show that $k' \Vdash C$. By definition, there is an n s.t. for all $k'' \succeq_n k'$, $k'' \Vdash A$ or $k'' \Vdash B$. In both cases it follows that $k'' \Vdash C$, since $k' \Vdash A \to C$ and $k' \Vdash B \to C$. By the covering lemma, $k' \Vdash C$.

Case $\exists^+ \colon A \to \exists x A$. Show that $k \Vdash (A \to \exists x A)[\eta]$. Assume that $k' \succeq k$ and $k' \Vdash A[\eta]$. Show that $k' \Vdash (\exists x A)[\eta]$. Since $\eta = \eta_x^{\eta(x)}$ there is an $a \in |\mathcal{B}|$ (namely $a := \eta(x)$) such that $k' \Vdash A[\eta_x^a]$. Hence, $k' \Vdash (\exists x A)[\eta]$.

Case $\exists^- \colon \exists x A \to (\forall x.A \to B) \to B$ and $x \notin \mathsf{FV}(B)$. We show that $k \Vdash (\exists x A \to (\forall x.A \to B) \to B)[\eta]$. Assume that $k' \succeq k$ and $k' \Vdash (\exists x A)[\eta]$ and $k' \Vdash (\forall x.A \to B)[\eta]$. We show $k' \Vdash B[\eta]$. By definition, there is an n such that for all $k'' \succeq_n k'$ we have $a \in |\mathcal{B}|$ and $k'' \Vdash A[\eta_x^a]$. From $k' \Vdash (\forall x.A \to B)[\eta]$ follows that $k'' \Vdash B[\eta_x^a]$, and since $x \notin \mathsf{FV}(B)$, from the coincidence lemma, $k'' \Vdash B[\eta]$. Then, finally, by the covering lemma $k' \Vdash B[\eta]$.

Case $\to^+$. Let $k \Vdash \Gamma$ hold. We show that $k \Vdash A \to B$. Assume $k' \succeq k$ and $k' \Vdash A$. Our goal is $k' \Vdash B$. We have $k' \Vdash \Gamma \cup \{A\}$. Thus, $k' \Vdash B$ by IH.

Case $\to^-$. Let $k \Vdash \Gamma$ hold. The IH gives us $k \Vdash A \to B$ and $k \Vdash A$. Hence $k \Vdash B$.

Case $\forall^+$. Let $k \Vdash \Gamma[\eta]$ and $x \notin \mathsf{FV}(\Gamma)$ hold. Show that $k \Vdash (\forall x A)[\eta]$, i.e. $k \Vdash A[\eta_x^a]$ for an arbitrary $a \in |\mathcal{B}|$. We have

$$\begin{array}{ll} k \Vdash \Gamma[\eta_x^a] & \text{by the coincidence lemma, since } x \notin \mathsf{FV}(\Gamma) \\ k \Vdash A[\eta_x^a] & \text{by IH.} \end{array}$$

Case $\forall^-$. Let $k \Vdash \Gamma[\eta]$. We show that $k \Vdash A[x := t][\eta]$. We have

$$\begin{array}{ll} k \Vdash (\forall x A)[\eta] & \text{by IH} \\ k \Vdash A[\eta_x^{\eta(t)}] & \text{by definition} \\ k \Vdash A[x := t][\eta] & \text{by the substitution lemma.} \end{array}$$

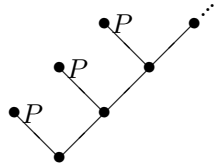
This concludes the proof. □

2.5. Counter Models. With soundness at hand, it is easy to build counter models for derivations not valid in minimal or intuitionistic logic.

A *Beth-structure* $\mathcal{B} = (D, I_0, I_1)$ *for intuitionistic logic* is a Beth-structure in which $\bot$ is never forced, i.e. $I_1(\bot, k) = 0$ for all k. Thus, in Beth-structures for intuitionistic logic we have

$$\begin{aligned} k \Vdash \neg A &\iff \forall k' \succeq k\, k' \nVdash A, \\ k \Vdash \neg\neg A &\iff \forall k' \succeq k\, k' \nVdash \neg A \\ &\iff \forall k' \succeq k \exists k'' \succeq k'\, k'' \Vdash A. \end{aligned}$$

As an example, we show that $\nvdash_i \neg\neg P \to P$. We describe the desired Beth-structure by means of a diagram below. Next to each node, we write the propositions forced on that node.



The it is easily seen that

$$\langle\rangle \nVdash P, \qquad \langle\rangle \Vdash \neg\neg P.$$

Thus $\langle\rangle \nVdash \neg\neg P \to P$ and hence $\nvdash \neg\neg P \to P$. Since for each R and all k, $k \Vdash \mathsf{Efq}_R$, it also follows that $\nvdash_i \neg\neg P \to P$. The model also shows that the Pierce formula $((P \to Q) \to P) \to P$ is invalid in intuitionistic logic.

3. Completeness of Minimal and Intuitionistic Logic

Next, we show the converse of soundness theorem, for minimal as well as intuitionistic logic.

3.1. Completeness of Minimal Logic.

THEOREM (Completeness). *Let $\Gamma \cup \{A\}$ be a set of formulas. Then the following propositions are equivalent.*

(a) $\Gamma \vdash A$.
(b) $\Gamma \Vdash A$, *i.e. for all Beth-structures $\mathcal{B}$, nodes k and assignments η*

$$\mathcal{B}, k \Vdash \Gamma[\eta] \Rightarrow \mathcal{B}, k \Vdash A[\eta].$$

PROOF. Soundness is one direction. For the other direction we employ a technique developed by Harvey Friedman and construct a Beth-structure $\mathcal{B}$ (over the set T_{01} of all finite 0-1-sequences k ordered by the initial segment relation $k \preceq k'$) with the property that $\Gamma \vdash B$ is equivalent to $\mathcal{B}, \langle\rangle \Vdash B[\mathsf{id}]$.

In order to define $\mathcal{B}$, we will need an enumeration $A_0, A_1, A_2, \ldots$ of $\mathcal{L}$-formulas, in which each formula occurs countably many times. We also fix an enumeration $x_0, x_1, \ldots$ of variables. Let $\Gamma = \bigcup_n \Gamma_n$ be the union of finite sets Γ_n such that $\Gamma_n \subseteq \Gamma_{n+1}$. With each node $k \in T_{01}$, we associate a finite set Δ_k of formulas by induction on the length of k.

Let $\Delta_{\langle\rangle} := \emptyset$. Take a node k such that $\mathsf{lh}(k) = n$ and suppose that Δ_k is already defined. Write $\Delta \vdash_n B$ to mean that there is a derivation of length $\leq n$ of B from Δ. We define Δ_{k0} and Δ_{k1} as follows:

Case 1. $\Gamma_n, \Delta_k \nvdash_n A_n$. Then let

$$\Delta_{k0} := \Delta_k \quad \text{and} \quad \Delta_{k1} := \Delta_k \cup \{A_n\}.$$

Case 2. $\Gamma_n, \Delta_k \vdash_n A_n = A'_n \vee A''_n$. Then let

$$\Delta_{k0} := \Delta_k \cup \{A_n, A'_n\} \quad \text{and} \quad \Delta_{k1} := \Delta_k \cup \{A_n, A''_n\}.$$

Case 3. $\Gamma_n, \Delta_k \vdash_n A_n = \exists x A'_n$. Then let

$$\Delta_{k0} := \Delta_{k1} := \Delta_k \cup \{A_n, A'_n[x := x_i]\}.$$

x_i is the first variable $\notin \mathsf{FV}(\Gamma_n, A_n, \Delta_k)$.

Case 4. $\Gamma_n, \Delta_k \vdash_n A_n$, and A_n is neither a disjunction nor an existentially quantified formula. Then let

$$\Delta_{k0} := \Delta_{k1} := \Delta_k \cup \{A_n\}.$$

Obviously $k \preceq k'$ implies that $\Delta_k \subseteq \Delta_{k'}$. We note that

$$\forall k' \succeq_n k \, \Gamma, \Delta_{k'} \vdash B \Rightarrow \Gamma, \Delta_k \vdash B. \tag{6}$$

It is sufficient to show

$$\Gamma, \Delta_{k0} \vdash B \quad \text{and} \quad \Gamma, \Delta_{k1} \vdash B \Rightarrow \Gamma, \Delta_k \vdash B.$$

In cases 1 and 4, this is obvious. For cases 2 and 3, it follows immediately from the axiom schemes $\vee^-$ and $\exists^-$.

Next, we show

$$\Gamma, \Delta_k \vdash B \Rightarrow \exists n \forall k' \succeq_n k \, B \in \Delta_{k'}. \tag{7}$$

We choose $n \geq \mathsf{lh}(k)$ such that $B = A_n$ and $\Gamma_n, \Delta_k \vdash_n A_n$. For all $k' \succeq k$, if $\mathsf{lh}(k') = n + 1$ then $A_n \in \Delta_{k'}$ (cf. the cases 2-4).

Using the sets Δ_k we can define an $\mathcal{L}$-Beth-structure $\mathcal{B}$ as $(\mathsf{Ter}_\mathcal{L}, I_0, I_1)$ (where $\mathsf{Ter}_\mathcal{L}$ denotes the set of terms of $\mathcal{L}$) and the canonical $I_0(f)\vec{t} := f\vec{t}$ and

$$\vec{t} \in I_1(R, k) \;\; :\Longleftrightarrow \;\; R\vec{t} \in \Delta_k.$$

Obviously, $t^\mathcal{B}[\mathsf{id}] = t$ for all $\mathcal{L}$-terms t.

We show that

$$\Gamma, \Delta_k \vdash B \iff \mathcal{B}, k \Vdash B[\mathsf{id}], \tag{8}$$

by induction on the complexity of B. For $\mathcal{B}, k \Vdash B[\mathsf{id}]$ we write $k \Vdash B$.

Case $R\vec{t}$. The following propositions are equivalent.

$\Gamma, \Delta_k \vdash R\vec{t}$	
$\exists n \forall k' \succeq_n k \, R\vec{t} \in \Delta_{k'}$	by (7) and (6)
$\exists n \forall k' \succeq_n k \, \vec{t} \in I_1(R, k')$	by definition of $\mathcal{B}$
$k \Vdash R\vec{t}$	by definition of $\Vdash$, since $t^\mathcal{B}[\mathsf{id}] = t$.

Case $B \vee C$. $\Rightarrow$. Let $\Gamma, \Delta_k \vdash B \vee C$. Choose an $n \geq \mathsf{lh}(k)$ such that $\Gamma_n, \Delta_k \vdash_n A_n = B \vee C$. Then, for all $k' \succeq k$ s.t. $\mathsf{lh}(k') = n$ it follows that

$$\Delta_{k'0} = \Delta_{k'} \cup \{B \vee C, B\} \quad \text{and} \quad \Delta_{k'1} = \Delta_{k'} \cup \{B \vee C, C\},$$

and by IH

$$k'0 \Vdash B \quad \text{and} \quad k'1 \Vdash C.$$

By definition, we have $k \Vdash B \vee C$. $\Leftarrow$.

$k \Vdash B \vee C$	
$\exists n \forall k' \succeq_n k \, . k' \Vdash B$ or $k' \Vdash C$	

$$\begin{array}{ll}
\exists n \forall k' \succeq_n k \,.\, \Gamma, \Delta_{k'} \vdash B \;\text{ or }\; \Gamma, \Delta_{k'} \vdash C & \text{by IH} \\
\exists n \forall k' \succeq_n k \, \Gamma, \Delta_{k'} \vdash B \vee C & \\
\Gamma, \Delta_k \vdash B \vee C & \text{by (6).}
\end{array}$$

The case $B \wedge C$ is evident.

Case $B \to C$. $\Rightarrow$. Let $\Gamma, \Delta_k \vdash B \to C$. We must show $k \Vdash B \to C$, i.e.,

$$\forall k' \succeq k . k' \Vdash B \Rightarrow k' \Vdash C.$$

Let $k' \succeq k$ be such that $k' \Vdash B$. By IH, it follows that $\Gamma, \Delta_{k'} \vdash B$, and $\Gamma, \Delta_{k'} \vdash C$ follows by assumption. Then again by IH $k' \Vdash C$.

$\Leftarrow$. Let $k \Vdash B \to C$, i.e. $\forall k' \succeq k . k' \Vdash B \Rightarrow k' \Vdash C$. We show that $\Gamma, \Delta_k \vdash B \to C$. At this point, we apply (6). Choose an $n \geq \mathsf{lh}(k)$ such that $B = A_n$. Let $k' \succeq_m k$ be such that $m := n - \mathsf{lh}(k)$. We show that $\Gamma, \Delta_{k'} \vdash B \to C$. If $\Gamma, \Delta_{k'} \vdash_n A_n$, then $k' \Vdash B$ by IH, and $k' \Vdash C$ by assumption, hence $\Gamma, \Delta_{k'} \vdash C$ again by IH and thus $\Gamma, \Delta_{k'} \vdash B \to C$.

If $\Gamma, \Delta_{k'} \nvdash_n A_n$ then by definition $\Delta_{k'1} = \Delta_{k'} \cup \{B\}$, hence $\Gamma, \Delta_{k'1} \vdash B$, and $k'1 \Vdash B$ by IH. Now $k'1 \Vdash C$ by assumption, and finally $\Gamma, \Delta_{k'1} \vdash C$ by IH. From $\Delta_{k'1} = \Delta_{k'} \cup \{B\}$, it follows that $\Gamma, \Delta_{k'} \vdash B \to C$.

Case $\forall x B$. The following propositions are equivalent.

$$\begin{array}{ll}
\Gamma, \Delta_k \vdash \forall x B & \\
\forall t \in \mathsf{Ter}_{\mathcal{L}} \, \Gamma, \Delta_k \vdash B[x := t] & \\
\forall t \in \mathsf{Ter}_{\mathcal{L}} \, k \Vdash B[x := t] & \text{by IH} \\
\forall t \in \mathsf{Ter}_{\mathcal{L}} \, k \Vdash B[\mathsf{id}^t_x] & \text{by the substitution lemma, since } t^{\mathcal{B}}[\mathsf{id}] = t \\
k \Vdash \forall x B & \text{by definition of } \Vdash.
\end{array}$$

Case $\exists x B$. This case is similar to the case $\vee$. The proof proceeds as follows. $\Rightarrow$. Let $\Gamma, \Delta_k \vdash \exists x B$. Choose an $n \geq \mathsf{lh}(k)$ such that $\Gamma_n, \Delta_k \vdash_n A_n = \exists x B$. Then, for all $k' \succeq k$ such that $\mathsf{lh}(k') = n$ it follows that

$$\Delta_{k'0} = \Delta_{k'1} = \Delta_k \cup \{\exists x B, B[x := x_i]\}$$

where x_i is not free in $\Delta_k \cup \{\exists x B\}$. Hence by IH

$$k'0 \Vdash B[x := x_i] \quad \text{and} \quad k'1 \Vdash B[x := x_i].$$

It follows by definition that $k \Vdash \exists x B$. $\Leftarrow$.

$$\begin{array}{ll}
k \Vdash \exists x B & \\
\exists n \forall k' \succeq_n k \exists t \in \mathsf{Ter}_{\mathcal{L}} \, k' \Vdash B[\mathsf{id}^t_x] & \\
\exists n \forall k' \succeq_n k \exists t \in \mathsf{Ter}_{\mathcal{L}} \, k' \Vdash B[x := t] & \\
\exists n \forall k' \succeq_n k \exists t \in \mathsf{Ter}_{\mathcal{L}} \, \Gamma, \Delta_{k'} \vdash B[x := t] & \text{by IH} \\
\exists n \forall k' \succeq_n k \, \Gamma, \Delta_{k'} \vdash \exists x B & \\
\Gamma, \Delta_k \vdash \exists x B & \text{by (6).}
\end{array}$$

Now, we are in a position to finalize the proof of the completeness theorem. We apply (b) to the Beth-structure $\mathcal{B}$ constructed above from Γ, the empty node $\langle\rangle$ and the assignment $\eta = \mathsf{id}$. Then $\mathcal{B}, \langle\rangle \Vdash \Gamma[\mathsf{id}]$ by (8), hence $\mathcal{B}, \langle\rangle \Vdash A[\mathsf{id}]$ by assumption and therefore $\Gamma \vdash A$ by (8) again. $\square$

3.2. Completeness of Intuitionistic Logic. Completeness of intuitionistic logic follows as a corollary.

COROLLARY. *Let $\Gamma \cup \{A\}$ be a set of formulas. The following propositions are equivalent.*

(a) $\Gamma \vdash_i A$.
(b) $\Gamma, \mathsf{Efq} \Vdash A$, *i.e., for all Beth-structures $\mathcal{B}$ for the intuitionistic logic, nodes k and assignments η*

$$\mathcal{B}, k \Vdash \Gamma[\eta] \Rightarrow \mathcal{B}, k \Vdash A[\eta]. \qquad \square$$

4. Completeness of Classical Logic

We give a proof of completeness of classical logic relying on the completeness proof for minimal logic above. Write $\Gamma \models A$ to mean that, for all structures $\mathcal{M}$ and assignments η,

$$\mathcal{M} \models \Gamma[\eta] \Rightarrow \mathcal{M} \models A[\eta].$$

4.1. The Completeness Theorem.

THEOREM (Completeness). *Let $\Gamma \cup \{A\}$ be a set of formulas (in $\mathcal{L}$). The following propositions are equivalent.*

(a) $\Gamma \vdash_c A$.
(b) $\Gamma \models A$.

PROOF. Soundness is one direction. For the other direction, we adapt the completeness of minimal logic.

Evidently, it is sufficient to treat formulas without $\vee$, $\exists$ and $\wedge$ (by Lemma 2.4).

Let $\Gamma \nvdash_c A$, i.e., $\Gamma, \mathsf{Stab} \nvdash A$. By the completeness theorem of minimal logic, there is a Beth-structure $\mathcal{B} = (\mathsf{Ter}_{\mathcal{L}}, I_0, I_1)$ on the complete binary tree T_{01} and a node l_0 such that $l_0 \Vdash \Gamma, \mathsf{Stab}$ and $l_0 \not\Vdash A$ (we write $k \Vdash B$ for $\mathcal{B}, k \Vdash B[\mathsf{id}]$).

A node k is *consistent* if $k \not\Vdash \bot$, and *stable* if $k \Vdash \mathsf{Stab}$. Let k be a stable node, and B a formula (without $\vee$, $\exists$ and $\wedge$). Then, $\mathsf{Stab} \vdash \neg\neg B \to B$ by the stability lemma. Hence, $k \Vdash \neg\neg B \to B$, and

$$\begin{aligned} k \not\Vdash B &\iff k \not\Vdash \neg\neg B \\ &\iff \exists k' \succeq k.\, k' \text{ consistent and } k' \Vdash \neg B. \end{aligned} \tag{9}$$

Let α be a branch in the underlying tree T_{01}. We define

$$\begin{aligned} \alpha \Vdash A &:\iff \exists k \in \alpha\, k \Vdash A, \\ \alpha \text{ is consistent} &:\iff \alpha \not\Vdash \bot, \\ \alpha \text{ is stable} &:\iff \exists k \in \alpha\, k \Vdash \mathsf{Stab}. \end{aligned}$$

Note that

$$\text{from } \alpha \Vdash \vec{A} \text{ and } \vdash \vec{A} \to B \text{ it follows that } \alpha \Vdash B. \tag{10}$$

To see this, consider $\alpha \Vdash \vec{A}$. Then $k \Vdash \vec{A}$ for a $k \in \alpha$, since α is linearly ordered. From $\vdash \vec{A} \to B$ it follows that $k \Vdash B$, i.e., $\alpha \Vdash B$.

A branch α is *generic* (in the sense that it generates a classical model) if it is consistent and stable, if in addition for all formulas B

$$\alpha \Vdash B \text{ or } \alpha \Vdash \neg B, \tag{11}$$

and for all formulas $\forall \vec{y} B$ (with $\vec{y}$ not empty) where B is not universally quantified

$$\forall \vec{s} \in \mathsf{Ter}_{\mathcal{L}}\, \alpha \Vdash B[\vec{y} := \vec{s}] \Rightarrow \alpha \Vdash \forall \vec{y} B \tag{12}$$

For a branch α, we define a classical structure $\mathcal{M}^\alpha = (\mathsf{Ter}_{\mathcal{L}}, I_0, I_1^\alpha)$ as

$$I_1^\alpha(R) := \bigcup_{k \in \alpha} I_1(R, k) \quad \text{for } R \neq \bot.$$

We show that for every generic branch α and each formula B with all connectives in $\{\rightarrow, \forall\}$

$$\alpha \Vdash B \iff \mathcal{M}^\alpha \models B. \tag{13}$$

The proof is by induction on the logical complexity of B.

Case $R\vec{t}$, $R \neq \bot$. Then the proposition holds for all α.

Case $\bot$. We have $\alpha \nVdash \bot$ for all consistent α.

Case $B \rightarrow C$. $\Rightarrow$. Let $\alpha \Vdash B \rightarrow C$ and $\mathcal{M}^\alpha \models B$. We must show that $\mathcal{M}^\alpha \models C$. Note that $\alpha \Vdash B$ by IH, hence $\alpha \Vdash C$, hence $\mathcal{M}^\alpha \models C$ again by IH. $\Leftarrow$. Let $\mathcal{M}^\alpha \models B \rightarrow C$. If $\mathcal{M}^\alpha \models B$, then $\mathcal{M}^\alpha \models C$, hence $\alpha \Vdash C$ by IH and therefore $\alpha \Vdash B \rightarrow C$. If $\mathcal{M}^\alpha \not\models B$, then $\alpha \nVdash B$ by IH, hence $\alpha \Vdash \neg B$ by (11) and therefore $\alpha \Vdash B \rightarrow C$, since α is stable (and $\vdash (\neg\neg C \rightarrow C) \rightarrow \bot \rightarrow C$).

Case $\forall \vec{y} B$ ($\vec{y}$ not empty) where B is not universally quantified. The following propositions are equivalent.

$$\begin{aligned}
&\alpha \Vdash \forall \vec{y} B \\
&\forall \vec{s} \in \mathsf{Ter}_{\mathcal{L}}\, \alpha \Vdash B[\vec{y} := \vec{s}] && \text{by (12)} \\
&\forall \vec{s} \in \mathsf{Ter}_{\mathcal{L}}\, \mathcal{M}^\alpha \models B[\vec{y} := \vec{s}] && \text{by IH} \\
&\mathcal{M}^\alpha \models \forall \vec{y} B.
\end{aligned}$$

We show that for each consistent stable node k, there is a generic branch containing k. For the purposes of the proof, we let $A_0, A_1, \ldots$ be an enumeration of formulas. We define a sequence $k = k_0 \preceq k_1 \preceq k_2 \ldots$ of consistent stable nodes inductively. Let $k_0 := k$. Assume that k_n is defined. We write A_n in the form $\forall \vec{y} B$ ($\vec{y}$ possibly empty) and B is not a universal formula. In case $k_n \Vdash \forall \vec{y} B$ let $k_{n+1} := k_n$. Otherwise we have $k_n \nVdash B[\vec{y} := \vec{s}]$ for some $\vec{s}$, and by (9) there is a consistent node $k' \succeq k_n$ such that $k' \Vdash \neg B[\vec{y} := \vec{s}]$. Let $k_{n+1} := k'$. Since $k_n \preceq k_{n+1}$, the node k_{n+1} is stable.

Let $\alpha := \{\, l \mid \exists n\ l \preceq k_n \,\}$, hence $k \in \alpha$. We show that α is generic. Clearly α is consistent and stable. The propositions (11) and (12) can be proved simultaneously. Let $C = \forall \vec{y} B$, where B is not a universal formula, and choose n, $C = A_n$. In case $k_n \Vdash \forall \vec{y} B$ we are done. Otherwise we have $k_n \nVdash B[\vec{y} := \vec{s}]$ for some $\vec{s}$, and by construction $k_{n+1} \Vdash \neg B[\vec{y} := \vec{s}]$. For (11) we get $k_{n+1} \Vdash \neg \forall \vec{y} B$ (since $\vdash \forall \vec{y} B \rightarrow B[\vec{y} := \vec{s}]$), and (12) follows from the consistency of α.

We are now in a position to give a proof of completeness. Since $l_0 \nVdash A$ and l_0 is stable, (9) yields a consistent node $k \succeq l_0$ such that $k \Vdash \neg A$.

Evidently, k is stable as well. By the proof above, there is a generic branch α such that $k \in \alpha$. Since $k \Vdash \neg A$ it follows that $\alpha \Vdash \neg A$, hence $\mathcal{M}^\alpha \models \neg A$ by (13). Moreover, $\alpha \Vdash \Gamma$, and $\mathcal{M}^\alpha \models \Gamma$ follow by (13). Then, $\Gamma \not\models A$. □

4.2. Compactness, Löwenheim-Skolem Theorem. The completeness theorem has many important corollaries. We mention only two. A set Γ of $\mathcal{L}$-formulas is *consistent* if $\Gamma \not\vdash_c \bot$, and *satisfiable* if there is an $\mathcal{L}$-structure $\mathcal{M}$ and an assignment η in $|\mathcal{M}|$ such that $\mathcal{M} \models B[\eta]$ for all $B \in \Gamma$.

Corollary. *Let Γ be a set of $\mathcal{L}$-formulas.*

(a) *If Γ is consistent, then Γ is satisfiable.*
(b) *(Compactness theorem). If each finite subset of Γ is satisfiable, Γ is satisfiable.*

Proof. (a). From $\Gamma \not\vdash_c \bot$ we obtain $\Gamma \not\models \bot$ by the completeness theorem, and this implies satisfiability of Γ.

(b). Otherwise we have $\Gamma \models \bot$, hence $\Gamma \vdash_c \bot$ by the completeness theorem, hence also $\Gamma_0 \vdash_c \bot$ for a finite subset $\Gamma_0 \subseteq \Gamma$, and therefore $\Gamma_0 \models \bot$ contrary to our assumption that Γ_0 has a model. □

Corollary (Löwenheim and Skolem). *Let Γ be a set of $\mathcal{L}$-formulas (we assume that $\mathcal{L}$ is countable). If Γ is satisfiable, then Γ is satisfiable on an $\mathcal{L}$-structure with a countable carrier set.*

Proof. We make use of the proof of the completeness theorem with $A = \bot$. It either yields $\Gamma \vdash_c \bot$ (which is excluded by assumption), or else a model of $\Gamma \cup \{\neg\bot\}$, whose carrier set is the countable set $\mathsf{Ter}_\mathcal{L}$. □

5. Uncountable Languages

We give a second proof of the completeness theorem for classical logic, which works for uncountable languages as well. This proof makes use of the *axiom of choice* (in the form of *Zorns lemma*).

5.1. Filters and Ultrafilters. Let $M \neq \emptyset$ be a set. $F \subseteq \mathcal{P}(M)$ is called *filter* on M, if

(a) $M \in F$ and $\emptyset \notin F$;
(b) if $X \in F$ and $X \subseteq Y \subseteq M$, then $Y \in F$;
(c) $X, Y \in F$ entails $X \cap Y \in F$.

F is called *ultrafilter*, if for all $X \in \mathcal{P}(M)$

$$X \in F \text{ or } M \setminus X \in F.$$

The intuition here is that the elements X of a filter F are considered to be "big". For instance, for M infinite the set $F = \{\, X \subseteq M \mid M \setminus X \text{ finite} \,\}$ is a filter.

Lemma. *Suppose F is an ultrafilter and $X \cup Y \in F$. Then $X \in F$ or $Y \in F$.*

Proof. If both X and Y are not in F, then $M \setminus X$ and $M \setminus Y$ are in F, hence also $(M \setminus X) \cap (M \setminus Y)$, which is $M \setminus (X \cup Y)$. This contradicts the assumption $X \cup Y \in F$. □

Let $M \neq \emptyset$ be a set and $S \subseteq \mathcal{P}(M)$. S has the *finite intersection property*, if $X_1 \cap \cdots \cap X_n \neq \emptyset$ for all $X_1, \ldots, X_n \in S$ and all $n \in \mathbb{N}$.

LEMMA. *If S has the finite intersection property, then there exists a filter F on M such that $F \supseteq S$.*

PROOF. $F := \{ X \mid X \supseteq X_1 \cap \cdots \cap X_n \text{ for some } X_1, \ldots, X_n \in S \}$. □

LEMMA. *Let $M \neq \emptyset$ be a set and F a filter on M. Then there is an ultrafilter U on M such that $U \supseteq F$.*

PROOF. By Zorns lemma (which will be proved - from the axiom of choice - in Chapter 5), there is a maximal filter U with $F \subseteq U$. We claim that U is an ultrafilter. So let $X \subseteq M$ and assume $X \notin U$ and $M \setminus X \notin U$. Since U is maximal, $U \cup \{X\}$ cannot have the finite intersection property; hence there is a $Y \in U$ such that $Y \cap X = \emptyset$. Similary we obtain $Z \in U$ such that $Z \cap (M \setminus X) = \emptyset$. But then $Y \cap Z = \emptyset$, a contradiction. □

5.2. Products and Ultraproducts. Let $M \neq \emptyset$ be a set and $A_i \neq \emptyset$ sets for $i \in M$. Let

$$\prod_{i \in M} A_i := \{ \alpha \mid \alpha \text{ is a function, } \mathsf{dom}(\alpha) = M \text{ and } \alpha(i) \in A_i \text{ for all } i \in M \}.$$

Observe that, by the *axiom of choice*, $\prod_{i \in M} A_i \neq \emptyset$. We write $\alpha \in \prod_{i \in M} A_i$ as $\langle \alpha(i) \mid i \in M \rangle$.

Now let $M \neq \emptyset$ be a set, F a filter on M and $\mathcal{A}_i$ structures for $i \in M$. Then the *F-product structure* $\mathcal{A} = \prod^F_{i \in M} \mathcal{A}_i$ is defined by

(a) $|\mathcal{A}| := \prod_{i \in M} |\mathcal{A}_i|$ (notice that $|\mathcal{A}| \neq \emptyset$).
(b) for an n-ary relation symbol R and $\alpha_1, \ldots, \alpha_n \in |\mathcal{A}|$ let

$$R^{\mathcal{A}}(\alpha_1, \ldots, \alpha_n) \iff \{ i \in M \mid R^{\mathcal{A}_i}(\alpha_1(i), \ldots, \alpha_n(i)) \} \in F.$$

(c) for an n-ary function symbol f and $\alpha_1, \ldots, \alpha_n \in |\mathcal{A}|$ let

$$f^{\mathcal{A}}(\alpha_1, \ldots, \alpha_n) := \langle f^{\mathcal{A}_i}(\alpha_1(i), \ldots, \alpha_n(i)) \mid i \in M \rangle.$$

For an ultrafilter U we call $\mathcal{A} = \prod^U_{i \in M} \mathcal{A}_i$ the *U-ultraproduct* of the $\mathcal{A}_i$ for $i \in M$.

5.3. The Fundamental Theorem on Ultraproducts. The properties of ultrafilters correspond in a certain sense to the definition of the consequence relation $\models$. For example, for an ultrafilter U we have

$$\begin{aligned} \mathcal{M} \models (A \vee B)[\eta] &\iff \mathcal{M} \models A[\eta] \text{ or } \mathcal{M} \models B[\eta] \\ X \cup Y \in U &\iff X \in U \text{ or } Y \in U \end{aligned}$$

and

$$\begin{aligned} \mathcal{M} \models \neg A[\eta] &\iff \mathcal{M} \not\models A[\eta] \\ X \notin U &\iff M \setminus X \in U. \end{aligned}$$

This is the background of the following theorem.

THEOREM (Fundamental Theorem on Ultraproducts, Łoś 1955). *Let $\mathcal{A} = \prod_{i\in M}^{U} \mathcal{A}_i$ be an U-ultraproduct, A a formula and η an assignment in $|\mathcal{A}|$. Then we have*

$$\mathcal{A} \models A[\eta] \iff \{\, i \in M \mid \mathcal{A}_i \models A[\eta_i] \,\} \in U,$$

where η_i is the assignment induced by $\eta_i(x) = \eta(x)(i)$ for $i \in M$.

PROOF. We first prove a similar property for terms.

$$t^{\mathcal{A}}[\eta] = \langle\, t^{\mathcal{A}_i}[\eta_i] \mid i \in M \,\rangle. \tag{14}$$

The proof is by induction on t. For a variable the claim follows from the definition. **Case** $ft_1 \dots t_n$. For simplicity assume $n = 1$; so we consider ft. We obtain

$$\begin{aligned}(ft)^{\mathcal{A}}[\eta] &= f^{\mathcal{A}}(t^{\mathcal{A}}[\eta]) \\ &= f^{\mathcal{A}}(\langle\, t^{\mathcal{A}_i}[\eta_i] \mid i \in M \,\rangle) \quad \text{by IH} \\ &= \langle\, (ft)^{\mathcal{A}_i}[\eta_i] \mid i \in M \,\rangle.\end{aligned}$$

Case $Rt_1 \dots t_n$. For simplicity assume $n = 1$; so consider Rt. We obtain

$$\begin{aligned}\mathcal{A} \models Rt[\eta] &\iff R^{\mathcal{A}}(t^{\mathcal{A}}[\eta]) \\ &\iff \{\, i \in M \mid R^{\mathcal{A}_i}(t^{\mathcal{A}}[\eta](i)) \,\} \in U \\ &\iff \{\, i \in M \mid R^{\mathcal{A}_i}(t^{\mathcal{A}_i}[\eta_i]) \,\} \in U \quad \text{by (14)} \\ &\iff \{\, i \in M \mid \mathcal{A}_i \models Rt[\eta_i] \,\} \in U.\end{aligned}$$

Case $A \to B$.

$\mathcal{A} \models (A \to B)[\eta]$
$\iff$ if $\mathcal{A} \models A[\eta]$, then $\mathcal{A} \models B[\eta]$
$\iff$ if $\{\, i \in M \mid \mathcal{A}_i \models A[\eta_i] \,\} \in U$, then $\{\, i \in M \mid \mathcal{A}_i \models B[\eta_i] \,\} \in U$ by IH
$\iff$ $\{\, i \in M \mid \mathcal{A}_i \models A[\eta_i] \,\} \notin U$ or $\{\, i \in M \mid \mathcal{A}_i \models B[\eta_i] \,\} \in U$
$\iff$ $\{\, i \in M \mid \mathcal{A}_i \models \neg A[\eta_i] \,\} \in U$ or $\{\, i \in M \mid \mathcal{A}_i \models B[\eta_i] \,\} \in U$ for U is an ultrafilter
$\iff$ $\{\, i \in M \mid \mathcal{A}_i \models (A \to B)[\eta_i] \,\} \in U$.

Case $\forall x A$.

$\mathcal{A} \models (\forall x A)[\eta]$
$\iff$ for all $\alpha \in |\mathcal{A}|$, $\mathcal{A} \models A[\eta^{\alpha}_x]$
$\iff$ for all $\alpha \in |\mathcal{A}|$, $\{\, i \in M \mid \mathcal{A}_i \models A[(\eta_i)^{\alpha(i)}_x] \,\} \in U$ by IH
(15) $\iff$ $\{\, i \in M \mid$ for all $a \in |\mathcal{A}_i|$, $\mathcal{A}_i \models A[(\eta_i)^{a}_x] \,\} \in U$ see below
$\iff$ $\{\, i \in M \mid \mathcal{A}_i \models (\forall x A)[\eta_i] \,\} \in U$.

It remains to show (15). Let $X := \{\, i \in M \mid$ for all $a \in |\mathcal{A}_i|$, $\mathcal{A}_i \models A[(\eta_i)^{a}_x] \,\}$ and $Y_\alpha := \{\, i \in M \mid \mathcal{A}_i \models A[(\eta_i)^{\alpha(i)}_x] \,\}$ for $\alpha \in |\mathcal{A}|$.

$\Leftarrow$. Let $\alpha \in |\mathcal{A}|$ and $X \in U$. Clearly $X \subseteq Y_\alpha$, hence also $Y_\alpha \in U$.

$\Rightarrow$. Let $Y_\alpha \in U$ for all α. Assume $X \notin U$. Since U is an ultrafilter,

$$M \setminus X = \{\, i \in M \mid \text{there is an } a \in |\mathcal{A}_i| \text{ such that } \mathcal{A}_i \not\models A[(\eta_i)^a_x] \,\} \in U.$$

We choose by the axiom of choice an $\alpha_0 \in |\mathcal{A}|$ such that

$$\alpha_0(i) = \begin{cases} \text{some } a \in |\mathcal{A}_i| \text{ such that } \mathcal{A}_i \not\models A[(\eta_i)^a_x] & \text{if } i \in M \setminus X, \\ \text{an arbitrary} \in |\mathcal{A}_i| & \text{otherwise.} \end{cases}$$

Then $Y_{\alpha_0} \cap (M \setminus X) = \emptyset$, contradicting $Y_{\alpha_0}, M \setminus X \in U$. □

If we choose $\mathcal{A}_i = \mathcal{B}$ constant, then $\mathcal{A} = \prod^U_{i \in M} \mathcal{B}$ satisfies the same formulas as $\mathcal{B}$ (such structures will be called *elementary equivalent* in section 6; the notation is $\mathcal{A} \equiv \mathcal{B}$). $\prod^U_{i \in M} \mathcal{B}$ is called an *ultrapower* of $\mathcal{B}$.

5.4. General Compactness and Completeness.

Corollary (General Compactness Theorem). *Every finitely satisfiable set Γ of formulas is satisfiable.*

Proof. Let $M := \{\, i \subseteq \Gamma \mid i \text{ finite} \,\}$. For $i \in M$ let $\mathcal{A}_i$ be a model of i under the assignment η_i. For $A \in \Gamma$ let $Z_A := \{\, i \in M \mid A \in i \,\} = \{\, i \subseteq \Gamma \mid i \text{ finite and } A \in i \,\}$. Then $F := \{\, Z_A \mid A \in \Gamma \,\}$ has the finite intersection property (for $\{A_1, \ldots, A_n\} \in Z_{A_1} \cap \ldots Z_{A_n}$). By the lemmata in 5.1 there is an ultrafilter U on M such that $F \subseteq U$. We consider $\mathcal{A} := \prod^U_{i \in M} \mathcal{A}_i$ and the product assigment η such that $\eta(x)(i) := \eta_i(x)$, and show $\mathcal{A} \models \Gamma[\eta]$. So let $A \in \Gamma$. By the theorem it suffices to show $X_A := \{\, i \in M \mid \mathcal{A}_i \models A[\eta_i] \,\} \in U$. But this follows form $Z_A \subseteq X_A$ and $Z_A \in F \subseteq U$. □

An immediate consequence is that if $\Gamma \models A$, then there exists a finite subset $\Gamma' \subseteq \Gamma$ such that $\Gamma' \models A$.

For every set Γ of formulas let $L(\Gamma)$ be the set of all function and relation symbols occurring in Γ. If $\mathcal{L}$ is a sublanguage of $\mathcal{L}'$, $\mathcal{M}$ an $\mathcal{L}$-structure and $\mathcal{M}'$ an $\mathcal{L}'$-structure, then $\mathcal{M}'$ is called an *expansion* of $\mathcal{M}$ (and $\mathcal{M}$ a *reduct* of $\mathcal{M}'$), if $|\mathcal{M}| = |\mathcal{M}'|$, $f^{\mathcal{M}} = f^{\mathcal{M}'}$ for all function symbols and $R^{\mathcal{M}} = R^{\mathcal{M}'}$ for all relation symbols in the language $\mathcal{L}$. The (uniquely determined) $\mathcal{L}$-reduct of $\mathcal{M}'$ is denoted by $\mathcal{M}' \restriction \mathcal{L}$. If $\mathcal{M}'$ is an expansion of $\mathcal{M}$ and η an assignment in $|\mathcal{M}|$, then clearly $t^{\mathcal{M}}[\eta] = t^{\mathcal{M}'}[\eta]$ for every $\mathcal{L}$-term t and $\mathcal{M} \models A[\eta]$ iff $\mathcal{M}' \models A[\eta]$ for every $\mathcal{L}$-formula A. Hence the validity of $\Gamma \models A$ does not depend on the underlying language $\mathcal{L}$, as long as $L(\Gamma \cup \{A\}) \subseteq \mathcal{L}$ (or more precisely $\subseteq \mathsf{Fun}_{\mathcal{L}} \cup \mathsf{Rel}_{\mathcal{L}}$).

Corollary (General Completeness Theorem). *Let $\Gamma \cup \{A\}$ be a set of formulas, where the underlying language may be uncountable. Then*

$$\Gamma \vdash_c A \iff \Gamma \models A.$$

Proof. One direction again is the soundness theorem. For the converse we can assume (by the first remark above) that for some finite $\Gamma' \subseteq \Gamma$ we have $\Gamma' \models A$. But then we have $\Gamma' \models A$ in a countable language (by the second remark above). By the completeness theorem for countable languages we obtain $\Gamma' \vdash_c A$, hence also $\Gamma \vdash_c A$. □

6. Basics of Model Theory

In this section we will (as is common in model theory) also allow uncountable languages $\mathcal{L}$. As we have just seen, completeness as well as compactness hold for such languages as well.

6.1. Equality Axioms. We first consider *equality axioms*. So we assume in this section that our underlying language $\mathcal{L}$ contains a binary relation symbol $=$. The set $\mathsf{Eq}_\mathcal{L}$ of $\mathcal{L}$-equality axioms consists of (the universal closures of)

$$\begin{aligned}
&x = x \quad \text{(reflexivity)},\\
&x = y \to y = x \quad \text{(symmetry)},\\
&x = y \to y = z \to x = z \quad \text{(transitivity)},\\
&x_1 = y_1 \to \cdots \to x_n = y_n \to f x_1 \dots x_n = f y_1 \dots y_n,\\
&x_1 = y_1 \to \cdots \to x_n = y_n \to R x_1 \dots x_n \to R y_1 \dots y_n,
\end{aligned}$$

for all n-ary function symbols f and relation symbols R of the language $\mathcal{L}$.

Lemma (Equality). (a) $\mathsf{Eq}_\mathcal{L} \vdash t = s \to r[x := t] = r[x := s]$.
(b) $\mathsf{Eq}_\mathcal{L} \vdash t = s \to (A[x := t] \leftrightarrow A[x := s])$.

Proof. (a). Induction on r. (b). Induction on A; we only consider the case $\forall y A$. Then $(\forall y A)[x := r] = \forall y A[x := r]$, and by IH we have $\mathsf{Eq}_\mathcal{L} \vdash t = s \to A[x := t] \to A[x := s]$. This entails the claim. □

An $\mathcal{L}$-structure $\mathcal{M}$ satisfies the equality axioms iff $=^\mathcal{M}$ is a *congruence relation* (i.e., an equivalence relation compatible with the functions and relations of $\mathcal{M}$). In this section we assume that all $\mathcal{L}$-structures considered $\mathcal{M}$ satisfy the equality axioms. The coincidence lemma then also holds with $=^\mathcal{M}$ instead of $=$:

Lemma (Coincidence). *Let η and ξ be assignments in $|\mathcal{M}|$ such that $\mathsf{dom}(\eta) = \mathsf{dom}(\xi)$ and $\eta(x) =^\mathcal{M} \xi(x)$ for all $x \in \mathsf{dom}(\eta)$. Then*
(a) *$t^\mathcal{M}[\eta] =^\mathcal{M} t^\mathcal{M}[\xi]$ if $\mathsf{vars}(t) \subseteq \mathsf{dom}(\eta)$ and*
(b) *$\mathcal{M} \models A[\eta] \iff \mathcal{M} \models A[\xi]$ if $\mathsf{FV}(A) \subseteq \mathsf{dom}(\eta)$.*

Proof. Induction on t and A, respectively. □

6.2. Cardinality of Models. Let $\mathcal{M}/=^\mathcal{M}$ be the *quotient structure*, whose carrier set consists of congruence classes. We call a structure $\mathcal{M}$ *infinite* (*countable*, *of cardinality n*), if $\mathcal{M}/=^\mathcal{M}$ is infinite (countable, of cardinality n).

By an *axiom system* Γ we understand a set of closed formulas such that $\mathsf{Eq}_{L(\Gamma)} \subseteq \Gamma$. A *model* of an axiom system Γ is an $\mathcal{L}$-structure $\mathcal{M}$ such that $L(\Gamma) \subseteq \mathcal{L}$ and $\mathcal{M} \models \Gamma$. For sets Γ of closed formulas we write

$$\mathsf{Mod}_\mathcal{L}(\Gamma) := \{\, \mathcal{M} \mid \mathcal{M} \text{ is an } \mathcal{L}\text{-structure and } \mathcal{M} \models \Gamma \cup \mathsf{Eq}_\mathcal{L} \,\}.$$

Clearly Γ is satisfiable iff Γ has a model.

Theorem. *If an axiom system has arbitrarily large finite models, then it has an infinite model.*

PROOF. Let Γ be such an axiom system. Suppose $x_0, x_1, x_2, \dots$ are distinct variables and

$$\Gamma' := \Gamma \cup \{\, x_i \neq x_j \mid i, j \in \mathbb{N} \text{ such that } i < j \,\}.$$

By assumption every finite subset of Γ' is satisfiable, hence by the general compactness theorem so is Γ'. Then we have $\mathcal{M}$ and η such that $\mathcal{M} \models \Gamma'[\eta]$ and therefore $\eta(x_i) \neq^{\mathcal{M}} \eta(x_j)$ for $i < j$. Hence $\mathcal{M}$ is infinite. □

6.3. Complete Theories, Elementary Equivalence. Let $\overline{\mathcal{L}}$ be the set of all closed $\mathcal{L}$-formulas. By a *theory* T we mean an axiom system closed under $\vdash_c$, so $\mathsf{Eq}_{L(T)} \subseteq T$ and

$$T = \{\, A \in \overline{L(T)} \mid T \vdash_c A \,\}.$$

A theory T is called *complete*, if for every formula $A \in \overline{L(T)}$, $T \vdash_c A$ or $T \vdash_c \neg A$.

For every $\mathcal{L}$-structure $\mathcal{M}$ (satisfying the equality axioms) the set of all closed $\mathcal{L}$-formulas A such that $\mathcal{M} \models A$ clearly is a theory; it is called the *theory of* $\mathcal{M}$ and denoted by $\mathsf{Th}(\mathcal{M})$.

Two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{M}'$ are called *elementarily equivalent* (written $\mathcal{M} \equiv \mathcal{M}'$), if $\mathsf{Th}(\mathcal{M}) = \mathsf{Th}(\mathcal{M}')$. Two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{M}'$ are called *isomorphic* (written $\mathcal{M} \cong \mathcal{M}'$), if there is a map $\pi\colon |\mathcal{M}| \to |\mathcal{M}'|$ inducing a bijection between $|\mathcal{M}/{=^{\mathcal{M}}}|$ and $|\mathcal{M}'/{=^{\mathcal{M}'}}|$, so

$$\forall a, b \in |\mathcal{M}|.\, a =^{\mathcal{M}} b \iff \pi(a) =^{\mathcal{M}'} \pi(b),$$
$$(\forall a' \in |\mathcal{M}'|)(\exists a \in |\mathcal{M}|)\, \pi(a) =^{\mathcal{M}'} a',$$

such that for all $a_1, \dots, a_n \in |\mathcal{M}|$

$$\pi(f^{\mathcal{M}}(a_1, \dots, a_n)) =^{\mathcal{M}'} f^{\mathcal{M}'}(\pi(a_1), \dots, \pi(a_n)),$$
$$R^{\mathcal{M}}(a_1, \dots, a_n) \iff R^{\mathcal{M}'}(\pi(a_1), \dots, \pi(a_n))$$

for all n-ary function symbols f and relation symbols R of the language $\mathcal{L}$.

We first collect some simple properties of the notions of the theory of a structure $\mathcal{M}$ and of elementary equivalence.

LEMMA. (a) $\mathsf{Th}(\mathcal{M})$ *ist complete.*

(b) *If* Γ *is an axiom system such that* $L(\Gamma) \subseteq \mathcal{L}$*, then*

$$\{\, A \in \overline{\mathcal{L}} \mid \Gamma \vdash_c A \,\} = \bigcap \{\, \mathsf{Th}(\mathcal{M}) \mid \mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(\Gamma) \,\}.$$

(c) $\mathcal{M} \equiv \mathcal{M}' \iff \mathcal{M} \models \mathsf{Th}(\mathcal{M}')$.

(d) *If* $\mathcal{L}$ *is countable, then for every* $\mathcal{L}$*-structure* $\mathcal{M}$ *there is a countable* $\mathcal{L}$*-structure* $\mathcal{M}'$ *such that* $\mathcal{M} \equiv \mathcal{M}'$.

PROOF. (a). Let $\mathcal{M}$ be an $\mathcal{L}$-structure and $A \in \overline{\mathcal{L}}$. Then $\mathcal{M} \models A$ or $\mathcal{M} \models \neg A$, hence $\mathsf{Th}(\mathcal{M}) \vdash_c A$ or $\mathsf{Th}(\mathcal{M}) \vdash_c \neg A$.

(b). For all $A \in \overline{\mathcal{L}}$ we have

$$\begin{aligned}
\Gamma \vdash_c A &\iff \Gamma \models A \\
&\iff \text{for all } \mathcal{L}\text{-structures } \mathcal{M},\ (\mathcal{M} \models \Gamma \Rightarrow \mathcal{M} \models A) \\
&\iff \text{for all } \mathcal{L}\text{-structures } \mathcal{M},\ (\mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(\Gamma) \Rightarrow A \in \mathsf{Th}(\mathcal{M})) \\
&\iff A \in \bigcap \{\, \mathsf{Th}(\mathcal{M}) \mid \mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(\Gamma) \,\}.
\end{aligned}$$

(c). $\Rightarrow$. Assume $\mathcal{M} \equiv \mathcal{M}'$ and $A \in \mathsf{Th}(\mathcal{M}')$. Then $\mathcal{M}' \models A$, hence $\mathcal{M} \models A$.

$\Leftarrow$. Assume $\mathcal{M} \models \mathsf{Th}(\mathcal{M}')$. Then clearly $\mathsf{Th}(\mathcal{M}') \subseteq \mathsf{Th}(\mathcal{M})$. For the converse inclusion let $A \in \mathsf{Th}(\mathcal{M})$. If $A \notin \mathsf{Th}(\mathcal{M}')$, by (a) we would also have $\neg A \in \mathsf{Th}(\mathcal{M}')$, hence $\mathcal{M} \models \neg A$ contradicting $A \in \mathsf{Th}(\mathcal{M})$.

(d). Let $\mathcal{L}$ be countable and $\mathcal{M}$ an $\mathcal{L}$-structure. Then $\mathsf{Th}(\mathcal{M})$ is satisfiable and therefore by the theorem of Löwenheim and Skolem possesses a satisfying $\mathcal{L}$-structure $\mathcal{M}'$ with a countable carrier set $\mathsf{Ter}_{\mathcal{L}}$. By (c), $\mathcal{M} \equiv \mathcal{M}'$. □

Moreover, we can characterize complete theories as follows:

THEOREM. *Let T be a theory and $\mathcal{L} = L(T)$. Then the following are equivalent.*

(a) *T is complete.*
(b) *For every model $\mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(T)$, $\mathsf{Th}(\mathcal{M}) = T$.*
(c) *Any two models $\mathcal{M}, \mathcal{M}' \in \mathsf{Mod}_{\mathcal{L}}(T)$ are elementarily equivalent.*

PROOF. (a) $\Rightarrow$ (b). Let T be complete and $\mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(T)$. Then $\mathcal{M} \models T$, hence $T \subseteq \mathsf{Th}(\mathcal{M})$. For the converse assume $A \in \mathsf{Th}(\mathcal{M})$. Then $\neg A \notin \mathsf{Th}(\mathcal{M})$, hence $\neg A \notin T$ and therefore $A \in T$.

(b) $\Rightarrow$ (c) is clear.

(c) $\Rightarrow$ (a). Let $A \in \overline{\mathcal{L}}$ and $T \nvdash_c A$. Then there is a model $\mathcal{M}_0$ of $T \cup \{\neg A\}$. Now let $\mathcal{M} \in \mathsf{Mod}_{\mathcal{L}}(T)$ be arbitrary. By (c) we have $\mathcal{M} \equiv \mathcal{M}_0$, hence $\mathcal{M} \models \neg A$. Therefore $T \vdash_c \neg A$. □

6.4. Elementary Equivalence and Isomorphism.

LEMMA. *Let π be an isomorphism between $\mathcal{M}$ and $\mathcal{M}'$. Then for all terms t and formulas A and for every sufficiently big assignment η in $|\mathcal{M}|$*

(a) *$\pi(t^{\mathcal{M}}[\eta]) =^{\mathcal{M}'} t^{\mathcal{M}'}[\pi \circ \eta]$ and*
(b) *$\mathcal{M} \models A[\eta] \iff \mathcal{M}' \models A[\pi \circ \eta]$. In particular,*

$$\mathcal{M} \cong \mathcal{M}' \Rightarrow \mathcal{M} \equiv \mathcal{M}'.$$

PROOF. (a). Induction on t. For simplicity we only consider the case of a unary function symbol.

$$\begin{aligned}
\pi(x^{\mathcal{M}}[\eta]) &= \pi(\eta(x)) = x^{\mathcal{M}'}[\pi \circ \eta] \\
\pi(c^{\mathcal{M}}[\eta]) &= \pi(c^{\mathcal{M}}) =^{\mathcal{M}'} c^{\mathcal{M}'} \\
\pi((ft)^{\mathcal{M}}[\eta]) &= \pi(f^{\mathcal{M}}(t^{\mathcal{M}}[\eta])) \\
&=^{\mathcal{M}'} f^{\mathcal{M}'}(\pi(t^{\mathcal{M}}[\eta])) \\
&=^{\mathcal{M}'} f^{\mathcal{M}'}(t^{\mathcal{M}'}[\pi \circ \eta]) \\
&= (ft)^{\mathcal{M}'}[\pi \circ \eta].
\end{aligned}$$

(b). Induction on A. For simplicity we only consider the case of a unary relation symbol and the case $\forall x A$.

$$\begin{aligned}
\mathcal{M} \models Rt[\eta] &\iff R^{\mathcal{M}}(t^{\mathcal{M}}[\eta]) \\
&\iff R^{\mathcal{M}'}(\pi(t^{\mathcal{M}}[\eta])) \\
&\iff R^{\mathcal{M}'}(t^{\mathcal{M}'}[\pi \circ \eta])
\end{aligned}$$

$$\iff \mathcal{M}' \models Rt[\pi \circ \eta],$$

$$\begin{aligned}
\mathcal{M} \models \forall x A[\eta] &\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M} \models A[\eta^a_x] \\
&\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M}' \models A[\pi \circ \eta^a_x] \\
&\iff \text{for all } a \in |\mathcal{M}|,\ \mathcal{M}' \models A[(\pi \circ \eta)^{\pi(a)}_x] \\
&\iff \text{for all } a' \in |\mathcal{M}'|,\ \mathcal{M}' \models A[(\pi \circ \eta)^{a'}_x] \\
&\iff \mathcal{M}' \models \forall x A[\pi \circ \eta]
\end{aligned}$$

This concludes the proof. □

The converse, i.e. that $\mathcal{M} \equiv \mathcal{M}'$ implies $\mathcal{M} \cong \mathcal{M}'$, is true for finite structures (see exercise sheet 9), but not for infinite ones:

THEOREM. *For every infinite structure $\mathcal{M}$ there is an elementarily equivalent structure $\mathcal{M}_0$ not isomorphic to $\mathcal{M}$.*

PROOF. Let $=^{\mathcal{M}}$ be the equality on $M := |\mathcal{M}|$, and let $\mathcal{P}(M)$ denote the power set of M. For every $\alpha \in \mathcal{P}(M)$ choose a new constant c_α. In the language $\mathcal{L}' := \mathcal{L} \cup \{ c_\alpha \mid \alpha \in \mathcal{P}(M) \}$ we consider the axiom system

$$\Gamma := \mathsf{Th}(\mathcal{M}) \cup \{ c_\alpha \neq c_\beta \mid \alpha, \beta \in \mathcal{P}(M) \text{ and } \alpha \neq \beta \} \cup \mathsf{Eq}_{\mathcal{L}'}.$$

Every finite subset of Γ is satisfiable by an appropriate expansion of $\mathcal{M}$. Hence by the general compactness theorem also Γ is satisfiable, say by $\mathcal{M}'_0$. Let $\mathcal{M}_0 := \mathcal{M}'_0 \restriction \mathcal{L}$. We may assume that $=^{\mathcal{M}_0}$ is the equality on $|\mathcal{M}_0|$. $\mathcal{M}_0$ is not isomorphic to $\mathcal{M}$, for otherwise we would have an injection of $\mathcal{P}(M)$ into M and therefore a contradiction. □

6.5. Non Standard Models. By what we just proved it is impossible to characterize an infinite structure by a first order axiom system up to isomorphism. However, if we extend first order logic by also allowing quantification over sets X, we can formulate the following *Peano axioms*

$$\begin{aligned}
&\forall n\, \mathsf{S}(n) \neq 0, \\
&\forall n \forall m. \mathsf{S}(n) = \mathsf{S}(m) \to n = m, \\
&\forall X. 0 \in X \to (\forall n. n \in X \to \mathsf{S}(n) \in X) \to \forall n\, n \in X.
\end{aligned}$$

One can show easily that $(\mathbb{N}, 0, \mathsf{S})$ is up to isomorphism the unique model of the Peano axioms. A structure which is elementarily equivalent, but not isomorphic to $\mathcal{N} := (\mathbb{N}, 0, \mathsf{S})$, is called a *non standard model* of the natural numbers. In non standard models of the natural numbers the principle of complete induction does not hold for all sets $X \subseteq \mathbb{N}$.

Similarly, a structure which is elementarily equivalent, but not isomorphic to $(\mathbb{R}, 0, 1, +, \cdot, <)$ is called a non standard model of the reals. In every non standard model of the reals the completeness axiom

$$\forall X. \emptyset \neq X \text{ bounded} \to \exists y. y = \sup(X)$$

does not hold for all sets $X \subseteq \mathbb{R}$.

THEOREM. *There are countable non standard models of the natural numbers.*

PROOF. Let x be a variable and

$$\Gamma := \mathsf{Th}(\mathcal{N}) \cup \{\, x \neq \underline{n} \mid n \in \mathbb{N} \,\},$$

where $\underline{0} := 0$ and $\underline{n+1} := \mathsf{S}\underline{n}$. Clearly every finite subset of Γ is satisfiable, hence by compactness also Γ. By the theorem of Löwenheim and Skolem we then have a countable or finite $\mathcal{M}$ and an assignment η such that $\mathcal{M} \models \Gamma[\eta]$. Because of $\mathcal{M} \models \mathsf{Th}(\mathcal{N})$ we have $\mathcal{M} \equiv \mathcal{N}$ by 6.3; hence $\mathcal{M}$ is countable. Moreover $\eta(x) \neq^{\mathcal{M}} \underline{n}^{\mathcal{M}}$ for all $n \in \mathbb{N}$, hence $\mathcal{M} \ncong \mathcal{N}$. □

6.6. Archimedian Ordered Fields. We now consider some easy applications to well-known axiom systems.

The axioms of *field theory* are (the equality axioms and)

$$\begin{aligned} &x + (y + z) = (x + y) + z, && x \cdot (y \cdot z) = (x \cdot y) \cdot z,\\ &0 + x = x, && 1 \cdot x = x,\\ &(-x) + x = 0, && x \neq 0 \to x^{-1} \cdot x = 1,\\ &x + y = y + x, && x \cdot y = y \cdot x, \end{aligned}$$

and also

$$\begin{aligned} &(x + y) \cdot z = (x \cdot z) + (y \cdot z),\\ &1 \neq 0. \end{aligned}$$

Fields are the models of this axiom system.

In the theory of *ordered fields* one has in addition a binary relation symbol $<$ and as axioms

$$\begin{aligned} &x \not< x,\\ &x < y \to y < z \to x < z,\\ &x < y \vee^{\mathsf{cl}} x = y \vee^{\mathsf{cl}} y < x,\\ &x < y \to x + z < y + z,\\ &0 < x \to 0 < y \to 0 < x \cdot y. \end{aligned}$$

Ordered fields are the models of this extended axiom system. An ordered field is called *archimedian ordered*, if for every element a of the field there is a natural number n such that a is less than the n-fold multiple of the 1 in the field.

THEOREM. *For every archimedian ordered field there is an elementarily equivalent ordered field that is not archimedian ordered.*

PROOF. Let $\mathcal{K}$ be an archimedian ordered field, x a variable and

$$\Gamma := \mathsf{Th}(\mathcal{K}) \cup \{\, \underline{n} < x \mid n \in \mathbb{N} \,\}.$$

Clearly every finite subset of Γ is satisfiable, hence by the general compactness theorem also Γ. Therefore we have $\mathcal{M}$ and η such that $\mathcal{M} \models \Gamma[\eta]$. Because of $\mathcal{M} \models \mathsf{Th}(\mathcal{K})$ we obtain $\mathcal{M} \equiv \mathcal{K}$ and hence $\mathcal{M}$ is an ordered field. Moreover $1^{\mathcal{M}} \cdot n <^{\mathcal{M}} \eta(x)$ for all $n \in \mathbb{N}$, hence $\mathcal{M}$ is not archimedian ordered. □

6.7. Axiomatizable Structures. A class $\mathcal{S}$ of $\mathcal{L}$-structures is called (*finitely*) *axiomatizable*, if there is a (finite) axiom system Γ such that $\mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\Gamma)$. Clearly $\mathcal{S}$ is finitely axiomatizable iff $\mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\{A\})$ for some formula A. If for every $\mathcal{M} \in \mathcal{S}$ there is an elementarily equivalent $\mathcal{M}' \notin \mathcal{S}$, then $\mathcal{S}$ cannot possibly be axiomatizable. By the theorem above we can conclude that the class of archimedian ordered fields is not axiomatizable. It also follows that the class of non archimedian ordered fields is not axiomatizable.

LEMMA. *Let $\mathcal{S}$ be a class of $\mathcal{L}$-structures and Γ an axiom system.*

(a) *$\mathcal{S}$ is finitely axiomatizable iff $\mathcal{S}$ and the complement of $\mathcal{S}$ are axiomatizable.*
(b) *If $\mathsf{Mod}_{\mathcal{L}}(\Gamma)$ is finitely axiomatizable, then there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\mathsf{Mod}_{\mathcal{L}}(\Gamma_0) = \mathsf{Mod}_{\mathcal{L}}(\Gamma)$.*

PROOF. (a). Let $1 - \mathcal{S}$ denote the complement of $\mathcal{S}$.

$\Rightarrow$. Let $\mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\{A\})$. Then $\mathcal{M} \in 1 - \mathcal{S} \iff \mathcal{M} \models \neg A$, hence $1 - \mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\{\neg A\})$.

$\Leftarrow$. Let $\mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\Gamma_1)$ and $1 - \mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\Gamma_2)$. Then $\Gamma_1 \cup \Gamma_2$ is not satisfiable, hence there is a finite $\Gamma \subseteq \Gamma_1$ such that $\Gamma \cup \Gamma_2$ is not satisfiable. One obtains

$$\mathcal{M} \in \mathcal{S} \Rightarrow \mathcal{M} \models \Gamma \Rightarrow \mathcal{M} \not\models \Gamma_2 \Rightarrow \mathcal{M} \notin 1 - \mathcal{S} \Rightarrow \mathcal{M} \in \mathcal{S},$$

hence $\mathcal{S} = \mathsf{Mod}_{\mathcal{L}}(\Gamma)$.

(b). Let $\mathsf{Mod}_{\mathcal{L}}(\Gamma) = \mathsf{Mod}_{\mathcal{L}}(\{A\})$. Then $\Gamma \models A$, hence also $\Gamma_0 \models A$ for a finite $\Gamma_0 \subseteq \Gamma$. One obtains

$$\mathcal{M} \models \Gamma \Rightarrow \mathcal{M} \models \Gamma_0 \Rightarrow \mathcal{M} \models A \Rightarrow \mathcal{M} \models \Gamma,$$

hence $\mathsf{Mod}_{\mathcal{L}}(\Gamma_0) = \mathsf{Mod}_{\mathcal{L}}(\Gamma)$. □

6.8. Complete Linear Orders Without End Points. Finally we consider as an example of a complete theory the theory DO of complete linear orders without end points. The axioms are (the equality axioms and)

$$\begin{aligned} &x \not< x, && x < y \to \exists^{\mathsf{cl}} z.x < z \wedge z < y, \\ &x < y \to y < z \to x < z, && \exists^{\mathsf{cl}} y\, x < y, \\ &x < y \vee^{\mathsf{cl}} x = y \vee^{\mathsf{cl}} y < x, && \exists^{\mathsf{cl}} y\, y < x. \end{aligned}$$

LEMMA. *Every countable model of DO is isomorphic to the structure $(\mathbb{Q}, <)$ of rational numbers.*

PROOF. Let $\mathcal{M} = (M, \prec)$ be a countable model of DO; we can assume that $=^{\mathcal{M}}$ is the equality on M. Let $M = \{\, b_n \mid n \in \mathbb{N} \,\}$ and $\mathbb{Q} = \{\, a_n \mid n \in \mathbb{N} \,\}$, where we may assume $a_n \neq a_m$ and $b_n \neq b_m$ for $n < m$. We define recursively functions $f_n \subseteq \mathbb{Q} \times M$ as follows. Let $f_0 := \{(a_0, b_0)\}$. Assume we have already constructed f_n.

Case $n+1 = 2m$. Let j be minimal such that $b_j \notin \mathsf{ran}(f_n)$. Choose $a_i \notin \mathsf{dom}(f_n)$ such that for all $a \in \mathsf{dom}(f_n)$ we have $a_i < a \leftrightarrow b_j < f_n(a)$; such an a_i exists, since $\mathcal{M}$ and $(\mathbb{Q}, <)$ are models of DO. Let $f_{n+1} := f_n \cup \{(a_i, b_j)\}$.

Case $n + 1 = 2m + 1$. This is treated similarly. Let i be minimal such that $a_i \notin \mathsf{dom}(f_n)$. Choose $b_j \notin \mathsf{ran}(f_n)$ such that for all $a \in \mathsf{dom}(f_n)$ we

have $a_i < a \leftrightarrow b_j < f_n(a)$; such a b_j exists, since $\mathcal{M}$ and $(\mathbb{Q}, <)$ are models of DO. Let $f_{n+1} := f_n \cup \{(a_i, b_j)\}$.

Then $\{b_0, \ldots, b_m\} \subseteq \mathsf{ran}(f_{2m})$ and $\{a_0, \ldots, a_{m+1}\} \subseteq \mathsf{dom}(f_{2m+1})$ by construction, and $f := \bigcup_n f_n$ is an isomorphism of $(\mathbb{Q}, <)$ onto $\mathcal{M}$. □

THEOREM. *The theory* DO *is complete, and* $\mathsf{DO} = \mathsf{Th}(\mathbb{Q}, <)$.

PROOF. Clearly $(\mathbb{Q}, <)$ is a model of DO. Hence by 6.3 it suffices to show that for every model $\mathcal{M}$ of DO we have $\mathcal{M} \equiv (\mathbb{Q}, <)$. So let $\mathcal{M}$ model of DO. By 6.3 there is a countable $\mathcal{M}'$ such that $\mathcal{M} \equiv \mathcal{M}'$. By the preceding lemma $\mathcal{M}' \cong (\mathbb{Q}, <)$, hence $\mathcal{M} \equiv \mathcal{M}' \equiv (\mathbb{Q}, <)$. □

A further example of a complete theory is the theory of algebraically closed fields. For a proof of this fact and for many more subjects of model theory we refer to the literature (e.g., the book of Chang and Keisler [**6**]).

7. Notes

The completeness theorem for classical logic has been proved by Gödel [**10**] in 1930. He did it for countable languages; the general case has been treated 1936 by Malzew [**17**]. Löwenheim and Skolem proved their theorem even before the completeness theorem was discovered: Löwenheim in 1915 [**16**] und Skolem in 1920 [**24**].

Beth-structures for intuitionistic logic have been introduced by Beth in 1956 [**1**]; however, the completeness proofs given there were in need of correction. 1959 Beth revised his paper in [**2**].

CHAPTER 3

Computability

In this chapter we develop the basics of recursive function theory, or as it is more generally known, computability theory. Its history goes back to the seminal works of Turing, Kleene and others in the 1930's.

A computable function is one defined by a program whose operational semantics tell an idealized computer what to do to its storage locations as it proceeds deterministically from input to output, without any prior restrictions on storage space or computation time. We shall be concerned with various program-styles and the relationships between them, but the emphasis throughout will be on one underlying data-type, namely the natural numbers, since it is there that the most basic foundational connections between proof theory and computation are to be seen in their clearest light.

The two best-known models of machine computation are the Turing Machine and the (Unlimited) Register Machine of Shepherdson and Sturgis [**22**]. We base our development on the latter since it affords the quickest route to the results we want to establish.

1. Register Machines

1.1. Programs. A *register machine* stores natural numbers in registers denoted u, v, w, x, y, z possibly with subscripts, and it responds step by step to a *program* consisting of an ordered list of basic instructions:

$$\begin{array}{l} I_0 \\ I_1 \\ \vdots \\ I_{k-1} \end{array}$$

Each instruction has one of the following three forms whose meanings are obvious:

Zero: $x := 0$
Succ: $x := x + 1$
Jump: **if** $x = y$ **then** I_m **else** I_n .

The instructions are obeyed in order starting with I_0 except when a conditional jump instruction is encountered, in which case the next instruction will be either I_m or I_n according as the numerical contents of registers x and y are equal or not at that stage. The computation *terminates* when it runs out of instructions, that is when the next instruction called for is I_k. Thus if a program of length k contains a jump instruction as above then it must satisfy the condition $m, n \leq k$ and I_k means "halt". Notice of course that some programs do not terminate, for example the following one-liner:

if $x = x$ **then** I_0 **else** I_1

1.2. Program Constructs. We develop some shorthand for building up standard sorts of programs.

Transfer. "$x := y$" is the program

$$\begin{array}{l} x := 0 \\ \textbf{if } x = y \textbf{ then } I_4 \textbf{ else } I_2 \\ x := x + 1 \\ \textbf{if } x = x \textbf{ then } I_1 \textbf{ else } I_1 \end{array}$$

which copies the contents of register y into register x.

Predecessor. The program "$x := y \dot{-} 1$" copies the modified predecessor of y into x, and simultaneously copies y into z:

$$\begin{array}{l} x := 0 \\ z := 0 \\ \textbf{if } x = y \textbf{ then } I_8 \textbf{ else } I_3 \\ z := z + 1 \\ \textbf{if } z = y \textbf{ then } I_8 \textbf{ else } I_5 \\ z := z + 1 \\ x := x + 1 \\ \textbf{if } z = y \textbf{ then } I_8 \textbf{ else } I_5\,. \end{array}$$

Composition. "P **;** Q" is the program obtained by concatenating program P with program Q. However in order to ensure that jump instructions in Q of the form "**if** $x = y$ **then** I_m **else** I_n" still operate properly within Q they need to be re-numbered by changing the addresses m, n to $k+m, k+n$ respectively where k is the length of program P. Thus the effect of this program is to do P until it halts (if ever) and then do Q.

Conditional. "**if** $x = y$ **then** P **else** Q **fi**" is the program

$$\begin{array}{l} \textbf{if } x = y \textbf{ then } I_1 \textbf{ else } I_{k+2} \\ \vdots P \\ \textbf{if } x = x \textbf{ then } I_{k+2+l} \textbf{ else } I_2 \\ \vdots Q \end{array}$$

where k, l are the lengths of the programs P, Q respectively, and again their jump instructions must be appropriately renumbered by adding 1 to the addresses in P and $k + 2$ to the addresses in Q. Clearly if $x = y$ then program P is obeyed and the next jump instruction automatically bipasses Q and halts. If $x \neq y$ then program Q is performed.

For Loop. "**for** $i = 1 \ldots x$ **do** P **od**" is the program

$$\begin{array}{l} i := 0 \\ \textbf{if } x = i \textbf{ then } I_{k+4} \textbf{ else } I_2 \\ i := i + 1 \\ \vdots P \\ \textbf{if } x = i \textbf{ then } I_{k+4} \textbf{ else } I_2 \end{array}$$

where again, k is the length of program P and the jump instructions in P must be appropriately re-addressed by adding 3. The intention of this new program is that it should iterate the program P x times (do nothing if $x = 0$). This requires the restriction that the register x and the "local" counting-register i are not re-assigned new values inside P.

While Loop. "**while** $x \neq 0$ **do** P **od**" is the program

$$\begin{array}{l}\textbf{if } x = 0 \textbf{ then } I_{k+2} \textbf{ else } I_1\\ \vdots P\\ \textbf{if } x = 0 \textbf{ then } I_{k+2} \textbf{ else } I_1\end{array}$$

where again, k is the length of program P and the jump instructions in P must be re-addressed by adding 1. This program keeps on doing P until (if ever) the register x becomes 0.

1.3. Computable Functions. A register machine program P may have certain distinguished "input registers" and "output registers". It may also use other "working registers" for scratchwork and these will initially be set to zero. We write $P(x_1, \ldots, x_k; y)$ to signify that program P has input registers $x_1, \ldots, x_k$ and one output register y, which are distinct.

DEFINITION. The program $P(x_1, \ldots, x_k; y)$ is said to *compute* the k-ary partial function $\varphi\colon \mathbb{N}^k \to \mathbb{N}$ if, starting with any numerical values $n_1, \ldots, n_k$ in the input registers, the program terminates with the number m in the output register if and only if $\varphi(n_1, \ldots, n_k)$ is defined with value m. In this case, the input registers hold their original values.

A function is *register machine computable* if there is some program which computes it.

Here are some examples.

Addition. "$\mathsf{Add}(x, y; z)$" is the program

$$z := x \textbf{ ; for } i = 1, \ldots, y \textbf{ do } z := z + 1 \textbf{ od}$$

which adds the contents of registers x and y into register z.

Subtraction. "$\mathsf{Subt}(x, y; z)$" is the program

$$z := x \textbf{ ; for } i = 1, \ldots, y \textbf{ do } w := z \dot{-} 1 \textbf{ ; } z := w \textbf{ od}$$

which computes the modified subtraction function $x \dot{-} y$.

Bounded Sum. If $P(x_1, \ldots, x_k, w; y)$ computes the $k+1$-ary function φ then the program $Q(x_1, \ldots, x_k, z; x)$:

$x := 0$ **;**
for $i = 1, \ldots, z$ **do** $w := i \dot{-} 1$ **;** $P(\vec{x}, w; y)$ **;** $v := x$ **;** $\mathsf{Add}(v, y; x)$ **od**

computes the function

$$\psi(x_1, \ldots, x_k, z) = \sum_{w<z} \varphi(x_1, \ldots, x_k, w)$$

which will be undefined if for some $w < z$, $\varphi(x_1, \ldots, x_k, w)$ is undefined.

Multiplication. Deleting "$w := i \dot{-} 1$ **;** P" from the last example gives a program $\mathsf{Mult}(z, y; x)$ which places the product of y and z into x.

Bounded Product. If in the bounded sum example, the instruction $x := x + 1$ is inserted immediately after $x := 0$, and if $\mathsf{Add}(v, y; x)$ is replaced by $\mathsf{Mult}(v, y; x)$, then the resulting program computes the function

$$\psi(x_1, \ldots, x_k, z) = \prod_{w<z} \varphi(x_1, \ldots, x_k, w)\,.$$

Composition. If $P_j(x_1, \ldots, x_k; y_j)$ computes φ_j for each $j = i, \ldots, m$ and if $P_0(y_1, \ldots, y_m; y_0)$ computes φ_0, then the program $Q(x_1, \ldots, x_k; y_0)$:

$$P_1(x_1, \ldots, x_k; y_1) \textbf{ ; } \ldots \textbf{ ; } P_m(x_1, \ldots, x_k; y_m) \textbf{ ; } P_0(y_1, \ldots, y_m; y_0)$$

computes the function

$$\psi(x_1, \ldots, x_k) = \varphi_0(\varphi_1(x_1, \ldots, x_k), \ldots, \varphi_m(x_1, \ldots, x_k))$$

which will be undefined if any of the φ-subterms on the right hand side is undefined.

Unbounded Minimization. If $P(x_1, \ldots, x_k, y; z)$ computes φ then the program $Q(x_1, \ldots, x_k; z)$:

$y := 0$ **;** $z := 0$ **;** $z := z + 1$ **;**
while $z \neq 0$ **do** $P(x_1, \ldots, x_k, y; z)$ **;** $y := y + 1$ **od ;**
$z := y \dot{-} 1$

computes the function

$$\psi(x_1, \ldots, x_k) \;=\; \mu y\,(\varphi(x_1, \ldots, x_k, y) = 0)$$

that is, the *least number* y such that $\varphi(x_1, \ldots, x_k, y')$ is defined for every $y' \leq y$ and $\varphi(x_1, \ldots, x_k, y) = 0$.

2. Elementary Functions

2.1. Definition and Simple Properties. The *elementary functions* of Kalmár (1943) are those number-theoretic functions which can be defined explicitly by compositional terms built up from variables and the constants $0, 1$ by repeated applications of addition $+$, modified subtraction $\dot{-}$, bounded sums and bounded products.

By omitting bounded products, one obtains the *subelementary* functions.

The examples in the previous section show that all elementary functions are computable and totally defined. Multiplication and exponentiation are elementary since

$$m \cdot n = \sum_{i<n} m \;\text{ and }\; m^n = \prod_{i<n} m$$

and hence by repeated composition, all exponential polynomials are elementary.

In addition the elementary functions are closed under

Definitions by Cases.

$$f(\vec{n}) = \begin{cases} g_0(\vec{n}) & \text{if } h(\vec{n}) = 0 \\ g_1(\vec{n}) & \text{otherwise} \end{cases}$$

since f can be defined from g_0, g_1 and h by

$$f(\vec{n}) = g_0(\vec{n}) \cdot (1 \dot{-} h(\vec{n})) + g_1(\vec{n}) \cdot (1 \dot{-} (1 \dot{-} h(\vec{n}))).$$

Bounded Minimization.

$$f(\vec{n}, m) = \mu\, k{<}m\, (g(\vec{n}, k) = 0)$$

since f can be defined from g by

$$f(\vec{n}, m) = \sum_{i<m} \big(1 \dot{-} \sum_{k\leq i} (1 \dot{-} g(\vec{n}, k))\big).$$

Note: this definition gives value m if there is no $k < m$ such that $g(\vec{n}, k) = 0$. It shows that not only the elementary, but in fact the subelementary functions are closed under bounded minimization. Furthermore, we define $\mu\, k{\leq}m\, (g(\vec{n}, k) = 0)$ as $\mu\, k{<}m{+}1\, (g(\vec{n}, k) = 0)$. Another notational convention will be that we shall often replace the brackets in $\mu\, k{<}m\, (g(\vec{n}, k) = 0)$ by a dot, thus: $\mu\, k{<}m.\, g(\vec{n}, k) = 0$.

LEMMA.

(a) *For every elementary function* $f\colon \mathbb{N}^r \to \mathbb{N}$ *there is a number* k *such that for all* $\vec{n} = n_1, \ldots, n_r$,

$$f(\vec{n}) \;<\; 2_k \max(\vec{n})$$

where $2_0(m) = m$ *and* $2_{k+1}(m) = 2^{2_k(m)}$.

(b) *Hence the function* $n \mapsto 2_n(1)$ *is not elementary.*

PROOF. (a). By induction on the build-up of the compositional term defining f. The result clearly holds if f is any one of the base functions:

$$f(\vec{n}) = 0 \text{ or } 1 \text{ or } n_i \text{ or } n_i + n_j \text{ or } n_i \dot{-} n_j.$$

If f is defined from g by application of bounded sum or product:

$$f(\vec{n}, m) = \sum_{i<m} g(\vec{n}, i) \text{ or } \prod_{i<m} g(\vec{n}, i)$$

where $g(\vec{n}, i) < 2_k \max(\vec{n}, i)$ then we have

$$f(\vec{n}, m) \leq 2_k \max(\vec{n}, m)^m < 2_{k+2} \max(\vec{n}, m)$$

(using $m^m < 2^{2^m}$). If f is defined from $g_0, g_1, \ldots, g_l$ by composition:

$$f(\vec{n}) = g_0(g_1(\vec{n}), \ldots, g_l(\vec{n}))$$

where for each $j \leq l$ we have $g_j(-) < 2_{k_j}(\max(-))$, then with $k = \max_j k_j$,

$$f(\vec{n}) < 2_k(2_k \max(\vec{n})) = 2_{2k} \max(\vec{n})$$

and this completes the first part.

(b). If $2_n(1)$ were an elementary function of n then by (a) there would be a positive k such that for all n,

$$2_n(1) < 2_k(n)$$

but then putting $n = 2_k(1)$ yields $2_{2_k(1)}(1) < 2_{2k}(1)$, a contradiction. □

2.2. Elementary Relations. A relation R on $\mathbb{N}^k$ is said to be *elementary* if its characteristic function

$$c_R(\vec{n}) = \begin{cases} 1 & \text{if } R(\vec{n}) \\ 0 & \text{otherwise} \end{cases}$$

is elementary. In particular, the "equality" and "less than" relations are elementary since their characteristic functions can be defined as follows:

$$c_<(m,n) = 1 \dot{-} (1 \dot{-} (n \dot{-} m)) \;\; ; \;\; c_=(m,n) = 1 \dot{-} (c_<(m,n) + c_<(n,m))) \, .$$

Furthermore if R is elementary then so is the function

$$f(\vec{n}, m) = \mu\, k{<}m\, R(\vec{n}, k)$$

since $R(\vec{n}, k)$ is equivalent to $1 \dot{-} c_R(\vec{n}, k) = 0$.

LEMMA. *The elementary relations are closed under applications of propositional connectives and bounded quantifiers.*

PROOF. For example, the characteristic function of $\neg R$ is

$$1 \dot{-} c_R(\vec{n}) \, .$$

The characteristic function of $R_0 \wedge R_1$ is

$$c_{R_0}(\vec{n}) \cdot c_{R_1}(\vec{n}).$$

The characteristic function of $\forall i{<}m\, R(\vec{n}, i)$ is

$$c_=(m, \mu i{<}m.\; c_R(\vec{n}, i) = 0).$$

□

EXAMPLES. The above closure properties enable us to show that many "natural" functions and relations of number theory are elementary; thus

$$\begin{aligned} \lfloor \frac{m}{n} \rfloor \quad &= \mu\, k{<}m\, (m < (k+1)n) \\ m \bmod n \; &= m \;\dot{-}\; \lfloor \frac{m}{n} \rfloor n \\ \mathsf{Prime}(m) &\leftrightarrow 1 < m \;\wedge\; \neg \exists n{<}m (1 < n \wedge m \bmod n = 0) \\ p_n \quad &= \mu\, m{<}2^{2^n} \left(\mathsf{Prime}(m) \;\wedge\; n = \sum_{i<m} c_{\mathsf{Prime}}(i) \right) \end{aligned}$$

so $p_0, p_1, p_2, \ldots$ gives the enumeration of primes in increasing order. The estimate $p_n \leq 2^{2^n}$ for the nth prime p_n can be proved by induction on n: For $n = 0$ this is clear, and for $n \geq 1$ we obtain

$$p_n \leq p_0 p_1 \cdots p_{n-1} + 1 \leq 2^{2^0} 2^{2^1} \cdots 2^{2^{n-1}} + 1 = 2^{2^n - 1} + 1 < 2^{2^n}.$$

2.3. The Class $\mathcal{E}$.

DEFINITION. The class $\mathcal{E}$ consists of those number theoretic functions which can be defined from the initial functions: constant 0, successor S, projections (onto the ith coordinate), addition $+$, modified subtraction $\dot{-}$, multiplication $\cdot$ and exponentiation 2^x, by applications of composition and bounded minimization.

The remarks above show immediately that the characteristic functions of the equality and less than relations lie in $\mathcal{E}$, and that (by the proof of the lemma) the relations in $\mathcal{E}$ are closed under propositional connectives and bounded quantifiers.

Furthermore the above examples show that all the functions in the class $\mathcal{E}$ are elementary. We now prove the converse, which will be useful later.

LEMMA. *There are "pairing functions" π, π_1, π_2 in $\mathcal{E}$ with the following properties:*

(a) π *maps* $\mathbb{N} \times \mathbb{N}$ *bijectively onto* $\mathbb{N}$,
(b) $\pi(a,b) < (a+b+1)^2$,
(c) $\pi_1(c), \pi_2(c) \leq c$,
(d) $\pi(\pi_1(c), \pi_2(c)) = c$,
(e) $\pi_1(\pi(a,b)) = a$,
(f) $\pi_2(\pi(a,b)) = b$.

PROOF. Enumerate the pairs of natural numbers as follows:

$$\begin{array}{cccccc} \vdots & & & & & \\ 10 & & & & & \\ 6 & \ldots & & & & \\ 3 & 7 & \ldots & & & \\ 1 & 4 & 8 & \ldots & & \\ 0 & 2 & 5 & 9 & \ldots & \end{array}$$

At position $(0,b)$ we clearly have the sum of the lengths of the preceeding diagonals, and on the next diagonal $a+b$ remains constant. Let $\pi(a,b)$ be the number written at position (a,b). Then we have

$$\pi(a,b) = \big(\sum_{i \leq a+b} i\big) + a = \frac{1}{2}(a+b)(a+b+1) + a.$$

Clearly $\pi\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is bijective. Moreover, $a, b \leq \pi(a,b)$ and in case $\pi(a,b) \neq 0$ also $a < \pi(a,b)$. Let

$$\pi_1(c) := \mu x{\leq}c \exists y{\leq}c\, (\pi(x,y) = c),$$
$$\pi_2(c) := \mu y{\leq}c \exists x{\leq}c\, (\pi(x,y) = c).$$

Then clearly $\pi_i(c) \leq c$ for $i \in \{1,2\}$ and

$$\pi_1(\pi(a,b)) = a,$$
$$\pi_2(\pi(a,b)) = b,$$
$$\pi(\pi_1(c), \pi_2(c)) = c.$$

π, π_1 and π_2 are elementary by definiton. □

REMARK. The above proof shows that π, π_1 and π_2 are in fact subelementary.

LEMMA (Gödel). *There is in $\mathcal{E}$ a function β with the following property: For every sequence $a_0, \ldots, a_{n-1} < b$ of numbers less than b we can find a number $c \leq 4 \cdot 4^{n(b+n+1)^4}$ such that $\beta(c,i) = a_i$ for all $i < n$.*

PROOF. Let

$$a := \pi(b,n) \quad \text{and} \quad d := \prod_{i<n}\big(1+\pi(a_i,i)a!\big).$$

From $a!$ and d we can, for each given $i<n$, reconstruct the number a_i as the unique $x<b$ such that

$$1+\pi(x,i)a! \mid d.$$

For clearly a_i is such an x, and if some $x<b$ were to satisfy the same condition, then because $\pi(x,i)<a$ and the numbers $1+ka!$ are relatively prime for $k\le a$, we would have $\pi(x,i)=\pi(a_j,j)$ for some $j<n$. Hence $x=a_j$ and $i=j$, thus $x=a_i$.

We can now define the *Gödel β-function* as

$$\beta(c,i) := \pi_1\big(\mu y{<}c.\ (1+\pi(\pi_1(y),i)\cdot\pi_1(c))\cdot\pi_2(y)=\pi_2(c)\big).$$

Clearly β is in $\mathcal{E}$. Furthermore with $c:=\pi(a!,d)$ we see that $\pi(a_i,\lceil d/1+\pi(a_i,i)a!\rceil)$ is the unique such y, and therefore $\beta(c,i)=a_i$. It is then not difficult to estimate the given bound on c, using $\pi(b,n)<(b+n+1)^2$. □

REMARK. The above definition of β shows that it is subelementary.

2.4. Closure Properties of $\mathcal{E}$.

THEOREM. *The class $\mathcal{E}$ is closed under limited recursion. Thus if g, h, k are given functions in $\mathcal{E}$ and f is defined from them according to the scheme*

$$\begin{aligned} f(\vec{m},0) &= g(\vec{m}) \\ f(\vec{m},n+1) &= h(n,f(\vec{m},n),\vec{m}) \\ f(\vec{m},n) &\le k(\vec{m},n) \end{aligned}$$

then f is in $\mathcal{E}$ also.

PROOF. Let f be defined from g, h and k in $\mathcal{E}$, by limited recursion as above. Using Gödel's β-function as in the last lemma we can find for any given $\vec{m},n$ a number c such that $\beta(c,i)=f(\vec{m},i)$ for all $i\le n$. Let $R(\vec{m},n,c)$ be the relation

$$\beta(c,0)=g(\vec{m}) \wedge \forall i{<}n.\ \beta(c,i+1)=h(i,\beta(c,i),\vec{m})$$

and note by the remarks above that its characteristic function is in $\mathcal{E}$. It is clear, by induction, that if $R(\vec{m},n,c)$ holds then $\beta(c,i)=f(\vec{m},i)$, for all $i\le n$. Therefore we can define f explicitly by the equation

$$f(\vec{m},n)=\beta(\mu c\ R(\vec{m},n,c),n).$$

f will lie in $\mathcal{E}$ if μc can be bounded by an $\mathcal{E}$ function. However, Lemma 2.3 gives a bound $4\cdot 4^{(n+1)(b+n+2)^4}$, where in this case b can be taken as the maximum of $k(\vec{m},i)$ for $i\le n$. But this can be defined in $\mathcal{E}$ as $k(\vec{m},i_0)$, where $i_0=\mu i{\le}n.\ \forall j{\le}n.\ k(\vec{m},j)\le k(\vec{m},i)$. Hence μc can be bounded by an $\mathcal{E}$ function. □

REMARK. Notice that it is in this proof only that the exponential function is required, in providing a bound for μ.

COROLLARY. *$\mathcal{E}$ is the class of all elementary functions.*

PROOF. It is sufficient merely to show that $\mathcal{E}$ is closed under bounded sums and bounded products. Suppose for instance, that f is defined from g in $\mathcal{E}$ by bounded summation: $f(\vec{m}, n) = \sum_{i<n} g(\vec{m}, i)$. Then f can be defined by limited recursion, as follows

$$\begin{aligned} f(\vec{m}, 0) &= 0 \\ f(\vec{m}, n+1) &= f(\vec{m}, n) + g(\vec{m}, n) \\ f(\vec{m}, n) &\leq n \cdot \max_{i<n} g(\vec{m}, i) \end{aligned}$$

and the functions (including the bound) from which it is defined are in $\mathcal{E}$. Thus f is in $\mathcal{E}$ by the last lemma. If instead, f is defined by bounded product, then proceed similarly. □

2.5. Coding Finite Lists. Computation on lists is a practical necessity, so because we are basing everything here on the single data type $\mathbb{N}$ we must develop some means of "coding" finite lists or sequences of natural numbers into $\mathbb{N}$ itself. There are various ways to do this and we shall adopt one of the most traditional, based on the pairing functions π, π_1, π_2.

The empty sequence is coded by the number 0 and a sequence n_0, n_1, $\ldots$, n_{k-1} is coded by the "sequence number"

$$\langle n_0, n_1, \ldots, n_{k-1} \rangle = \pi'(\ldots \pi'(\pi'(0, n_0), n_1), \ldots, n_{k-1})$$

with $\pi'(a, b) := \pi(a, b) + 1$, thus recursively,

$$\begin{aligned} \langle\rangle &:= 0, \\ \langle n_0, n_1, \ldots, n_k \rangle &:= \pi'(\langle n_0, n_1, \ldots, n_{k-1} \rangle, n_k). \end{aligned}$$

Because of the surjectivity of π, every number a can be decoded uniquely as a sequence number $a = \langle n_0, n_1, \ldots, n_{k-1} \rangle$. If a is greater than zero, $\mathsf{hd}(a) := \pi_2(a \dot{-} 1)$ is the "head" (i.e. rightmost element) and $\mathsf{tl}(a) := \pi_1(a \dot{-} 1)$ is the "tail" of the list. The kth iterate of tl is denoted $\mathsf{tl}^{(k)}$ and since $\mathsf{tl}(a)$ is less than or equal to a, $\mathsf{tl}^{(k)}(a)$ is elementarily definable (by limited recursion). Thus we can define elementarily the "length" and "decoding" functions:

$$\begin{aligned} \mathsf{lh}(a) &:= \mu k{\leq}a.\ \mathsf{tl}^{(k)}(a) = 0, \\ (a)_i &:= \mathsf{hd}(\mathsf{tl}^{(\mathsf{lh}(a) \dot{-} (i+1))}(a)). \end{aligned}$$

Then if $a = \langle n_0, n_1, \ldots, n_{k-1} \rangle$ it is easy to check that

$$\mathsf{lh}(a) = k \text{ and } (a)_i = n_i \text{ for each } i < k.$$

Furthermore $(a)_i = 0$ when $i \geq \mathsf{lh}(a)$. We shall write $(a)_{i,j}$ for $((a)_i)_j$ and $(a)_{i,j,k}$ for $(((a)_i)_j)_k$. This elementary coding machinery will be used at various crucial points in the following.

Note that our previous remarks show that the functions lh and $(a)_i$ are subelementary, and so is $\langle n_0, n_1, \ldots, n_{k-1} \rangle$ for each fixed k .

Concatenation of sequence numbers $b \star a$ is defined thus:

$$\begin{aligned} b \star \langle\rangle &:= b, \\ b \star \langle n_0, n_1, \ldots, n_k \rangle &:= \pi(b \star \langle n_0, n_1, \ldots, n_{k-1} \rangle, n_k) + 1. \end{aligned}$$

To check that this operation is also elementary, define $h(b,a,i)$ by recursion on i as follows.

$$\begin{aligned} h(b,a,0) &= b, \\ h(b,a,i+1) &= \pi(h(b,a,i),(a)_i)+1 \end{aligned}$$

and note that since $\pi(h(b,a,i),(a)_i) < (h(b,a,i)+a)^2$ it follows by induction on i that $h(b,a,i)$ is less than or equal to $(b+a+i)^{2^i}$. Thus h is definable by limited recursion from elementary functions and hence is itself elementary. Finally

$$b \star a = h(b,a,\mathsf{lh}(a)).$$

LEMMA. *The class $\mathcal{E}$ is closed under limited course-of-values recursion. Thus if h, k are given functions in $\mathcal{E}$ and f is defined from them according to the scheme*

$$\begin{aligned} f(\vec{m},n) &= h(n,\langle f(\vec{m},0),\ldots.f(\vec{m},n-1)\rangle,\vec{m}) \\ f(\vec{m},n) &\le k(\vec{m},n) \end{aligned}$$

then f is in $\mathcal{E}$ also.

PROOF. $\bar{f}(\vec{m},n) := \langle f(\vec{m},0),\ldots.f(\vec{m},n-1)\rangle$ is definable by

$$\begin{aligned} \bar{f}(\vec{m},0) &= 0, \\ \bar{f}(\vec{m},n+1) &= \bar{f}(\vec{m},n) \star \langle h(n,\bar{f}(\vec{m},n),\vec{m})\rangle \\ \bar{f}(\vec{m},n) &\le \Big(\sum_{i\le n} k(\vec{m},i)+1\Big)^{2^n}, \quad \text{using } \langle \underbrace{n,\ldots,n}_{k}\rangle < (n+1)^{2^k} \end{aligned}$$

□

3. The Normal Form Theorem

3.1. Program Numbers. The three types of register machine instructions I can be coded by "instruction numbers" $\sharp I$ thus, where $v_0, v_1, v_2, \ldots$ is a list of all variables used to denote registers:

If I is "$v_j := 0$" then $\sharp I = \langle 0,j\rangle$.
If I is "$v_j := v_j + 1$" then $\sharp I = \langle 1,j\rangle$.
If I is "**if** $v_j = v_l$ **then** I_m **else** I_n" then $\sharp I = \langle 2,j,l,m,n\rangle$.

Clearly, using the sequence coding and decoding apparatus above, we can check elementarily whether or not a given number is an instruction number.

Any register machine program $P = I_0, I_1, \ldots, I_{k-1}$ can then be coded by a "program number" or "index" $\sharp P$ thus:

$$\sharp P \;=\; \langle\, \sharp I_0, \sharp I_1, \ldots, \sharp I_{k-1}\,\rangle$$

and again (although it is tedious) we can elementarily check whether or not a given number is indeed of the form $\sharp P$ for some program P. Tradition has it that e is normally reserved as a variable over putative program numbers.

Standard program constructs such as those in Section 1 have associated "index-constructors", i.e. functions which, given indices of the subprograms, produce an index for the constructed program. The point is that for standard program constructs the associated index-constructor functions are elementary. For example there is an elementary index-constructor comp such

that, given programs P_0, P_1 with indices e_0, e_1, $\mathsf{comp}(e_0, e_1)$ is an index of the program P_0 ; P_1. A moment's thought should convince the reader that the appropriate definition of comp is as follows:

$$\mathsf{comp}(e_0, e_1) = e_0 \star \langle r(e_0, e_1, 0), r(e_0, e_1, 1), \ldots, r(e_0, e_1, \mathsf{lh}(e_1) \dot{-} 1)\rangle$$

where $r(e_0, e_1, i) =$

$$\begin{cases} \langle 2, (e_1)_{i,1}, (e_1)_{i,2}, (e_1)_{i,3} + \mathsf{lh}(e_0), (e_1)_{i,4} + \mathsf{lh}(e_0)\rangle & \text{if } (e_1)_{i,0} = 2 \\ (e_1)_i & \text{otherwise} \end{cases}$$

re-addresses the jump instructions in P_1. Clearly r and hence comp are elementary functions.

Definition. Henceforth, $\varphi_e^{(r)}$ denotes the partial function computed by the register machine program with program number e, operating on the input registers $v_1, \ldots, v_r$ and with output register v_0. There is no loss of generality here, since the variables in any program can always be renamed so that $v_1, \ldots, v_r$ become the input registers and v_0 the output. If e is not a program number, or it is but does not operate on the right variables, then we adopt the convention that $\varphi_e^{(r)}(n_1, \ldots, n_r)$ is undefined for all inputs $n_1, \ldots, n_r$.

3.2. Normal Form.

Theorem (Kleene's Normal Form). *For each arity r there is an elementary function U and an elementary relation T such that, for all e and all inputs $n_1, \ldots, n_r$,*

- $\varphi_e^{(r)}(n_1, \ldots, n_r)$ *is defined* $\iff \exists s\, T(e, n_1, \ldots, n_r, s)$
- $\varphi_e^{(r)}(n_1, \ldots, n_r) = U(\, e, n_1, \ldots, n_r, \mu\, s\, T(e, n_1, \ldots, n_r, s)\,)$.

Proof. A computation of a register machine program $P(v_1, \ldots, v_r; v_0)$ on numerical inputs $\vec{n} = n_1, \ldots, n_r$ proceeds deterministically, step by step, each step corresponding to the execution of one instruction. Let e be its program number, and let $v_0, \ldots, v_l$ be all the registers used by P, including the "working registers" so $r \leq l$.

The "state" of the computation at step s is defined to be the sequence number

$$\mathsf{state}(e, \vec{n}, s) = \langle e, i, m_0, m_1, \ldots, m_l\rangle$$

where $m_0, m_1, \ldots, m_l$ are the values stored in the registers $v_0, v_1, \ldots, v_l$ after step s is completed, and the next instruction to be performed is the ith one, thus $(e)_i$ is its instruction number.

The "state transition function" $\mathsf{tr}\colon \mathbb{N} \to \mathbb{N}$ computes the "next state". So suppose that $x = \langle e, i, m_0, m_1, \ldots, m_l\rangle$ is any putative state. Then in what follows, $e = (x)_0$, $i = (x)_1$, and $m_j = (x)_{j+2}$ for each $j \leq l$. The definition of $\mathsf{tr}(x)$ is therefore as follows:

$$\mathsf{tr}(x) = \langle e, i', m_0', m_1', \ldots, m_l'\rangle$$

where

- If $(e)_i = \langle 0, j\rangle$ where $j \leq l$ then $i' = i + 1$, $m_j' = 0$, and all other registers remain unchanged, i.e. $m_k' = m_k$ for $k \neq j$.

- If $(e)_i = \langle 1, j\rangle$ where $j \le l$ then $i' = i+1$, $m'_j = m_j + 1$, and all other registers remain unchanged.
- If $(e)_i = \langle 2, j_0, j_1, i_0, i_1\rangle$ where $j_0, j_1 \le l$ and $i_0, i_1 \le \mathsf{lh}(e)$ then $i' = i_0$ or $i' = i_1$ according as $m_{j_0} = m_{j_1}$ or not, and all registers remain unchanged, i.e. $m'_j = m_j$ for all $j \le l$.
- Otherwise, if x is not a sequence number, or if e is not a program number, or if it refers to a register v_k with $l < k$, or if $\mathsf{lh}(e) \le i$, then $\mathsf{tr}(x)$ simply repeats the same state x so $i' = i$, and $m'_j = m_j$ for every $j \le l$.

Clearly tr is an *elementary* function, since it is defined by elementarily decidable cases, with (a great deal of) elementary decoding and re-coding involved in each case.

Consequently, the "state function" $\mathsf{state}(e, \vec{n}, s)$ is also *elementary* because it can be defined by iterating the transition function by limited recursion on s as follows:

$$\begin{aligned}\mathsf{state}(e, \vec{n}, 0) &= \langle e, 0, n_1, \ldots, n_r, 0, \ldots, 0\rangle\\ \mathsf{state}(e, \vec{n}, s+1) &= \mathsf{tr}(\,\mathsf{state}(e, \vec{n}, s)\,)\\ \mathsf{state}(e, \vec{n}, s) &\le h(e, \vec{n}, s)\end{aligned}$$

where for the bounding function h we can take

$$h(e, \vec{n}, s) = \langle e, e\rangle \star \langle \max(\vec{n}) + s, \ldots, \max(\vec{n}) + s\rangle,$$

This is because the maximum value of any register at step s cannot be greater than $\max(\vec{n}) + s$. Now this expression clearly is elementary, since $\langle m, \ldots, m\rangle$ with i occurrences of m is definable by a limited recursion with bound $(m+i)^{2^i}$, as is easily seen by induction on i.

Now recall that if program P has program number e then computation terminates when instruction $I_{\mathsf{lh}(e)}$ is encountered. Thus we can define the "termination relation" $T(e, \vec{n}, s)$ meaning "computation terminates at step s", by

$$T(e, \vec{n}, s) \iff (\,\mathsf{state}(e, \vec{n}, s)\,)_1 = \mathsf{lh}(e).$$

Clearly T is elementary and

$$\varphi_e^{(r)}(\vec{n}) \text{ is defined } \iff \exists s\, T(e, \vec{n}, s).$$

The output on termination is the value of register v_0, so if we define the "output function" $U(e, \vec{n}, s)$ by

$$U(e, \vec{n}, s) \;=\; (\,\mathsf{state}(e, \vec{n}, s)\,)_2$$

then U is also elementary and

$$\varphi_e^{(r)}(\vec{n}) \;=\; U(e, \vec{n}, \mu s\, T(e, \vec{n}, s)).$$

This completes the proof. □

3.3. Σ^0_1-Definable Relations and μ-Recursive Functions. A relation R of arity r is said to be Σ^0_1-*definable* if there is an elementary relation E, say of arity $r + l$, such that for all $\vec{n} = n_1, \ldots, n_r$,

$$R(\vec{n}) \iff \exists k_1 \ldots \exists k_l\, E(\vec{n}, k_1, \ldots, k_l).$$

A partial function φ is said to be Σ^0_1-*definable* if its graph

$$\{\,(\vec{n}, m) \mid \varphi(\vec{n}) \text{ is defined and } = m\,\}$$

is Σ^0_1-definable.

To say that a non-empty relation R is Σ^0_1-definable is equivalent to saying that the set of all sequences $\langle \vec{n} \rangle$ satisfying R can be enumerated (possibly with repetitions) by some elementary function $f\colon \mathbb{N} \to \mathbb{N}$. Such relations are called *elementarily enumerable*. For choose any fixed sequence $\langle a_1, \ldots, a_r \rangle$ satisfying R and define

$$f(m) \;=\; \begin{cases} \langle (m)_1, \ldots, (m)_r \rangle & \text{if } E((m)_1, \ldots, (m)_{r+l}) \\ \langle a_1, \ldots, a_r \rangle & \text{otherwise.} \end{cases}$$

Conversely if R is elementarily enumerated by f then

$$R(\vec{n}) \iff \exists m\,(f(m) = \langle \vec{n} \rangle)$$

is a Σ^0_1-definition of R.

The *μ-recursive functions* are those (partial) functions which can be defined from the initial functions: constant 0, successor S, projections (onto the ith coordinate), addition $+$, modified subtraction $\dot{-}$ and multiplication $\cdot$, by applications of composition and unbounded minimization. Note that it is through unbounded minimization that partial functions may arise.

Lemma. *Every elementary function is μ-recursive.*

Proof. By simply removing the bounds on μ in the lemmas in 2.3 one obtains μ-recursive definitions of the pairing functions π, π_1, π_2 and of Gödel's β-function. Then by removing all mention of bounds from Theorem in 2.4 one sees that the μ-recursive functions are closed under (unlimited) primitive recursive definitions: $f(\vec{m}, 0) = g(\vec{m})$, $f(\vec{m}, n+1) = h(n, f(\vec{m}, n))$. Thus one can μ-recursively define bounded sums and bounded products, and hence all elementary functions. □

3.4. Computable Functions.

Definition. The *while-programs* are those programs which can be built up from assignment statements $x := 0$, $x := y$, $x := y + 1$, $x := y \dot{-} 1$, by Conditionals, Composition, For-Loops and While-Loops as in the subsection on program constructs in Section 1.

Theorem. *The following are equivalent:*

(a) *φ is register machine computable,*
(b) *φ is Σ^0_1-definable,*
(c) *φ is μ-recursive,*
(d) *φ is computable by a while program.*

Proof. The Normal Form Theorem shows immediately that every register machine computable function $\varphi^{(r)}_e$ is Σ^0_1-definable since

$$\varphi^{(r)}_e(\vec{n}) = m \iff \exists s.T(e, \vec{n}, s) \wedge U(e, \vec{n}, s) = m$$

and the relation $T(e, \vec{n}, s) \wedge U(e, \vec{n}, s) = m$ is clearly elementary. If φ is Σ^0_1-definable, say

$$\varphi(\vec{n}) = m \iff \exists k_1 \ldots \exists k_l\, E(\vec{n}, m, k_1, \ldots, k_l)$$

then φ can be defined μ-recursively by

$$\varphi(\vec{n}) \;=\; (\,\mu m\, E(\vec{n}, (m)_0, (m)_1, \ldots, (m)_l)\,)_0\ ,$$

using the fact (above) that elementary functions are μ-recursive. The examples of computable functionals in Section 1 show how the definition of any μ-recursive function translates automatically into a while program. Finally, the subsection on program constructs in Section 1 shows how to implement any while program on a register machine. □

Henceforth *computable* means "register machine computable" or any of its equivalents.

COROLLARY. *The function $\varphi_e^{(r)}(n_1, \ldots, n_r)$ is a computable partial function of the $r+1$ variables $e, n_1, \ldots, n_r$.*

PROOF. Immediate from the Normal Form. □

LEMMA. *A relation R is computable if and only if both R and its complement $\mathbb{N}^n \setminus R$ are Σ_1^0-definable.*

PROOF. We can assume that both R and $\mathbb{N}^n \setminus R$ are not empty, and (for simplicity) also $n = 1$.

$\Rightarrow$. By the theorem above every computable relation is Σ_1^0-definable, and with R clearly its complement is computable.

$\Leftarrow$. Let $f, g \in \mathcal{E}$ enumerate R and $\mathbb{N} \setminus R$, respectively. Then

$$h(n) := \mu i.f(i) = n \vee g(i) = n$$

is a total μ-recursive function, and $R(n) \leftrightarrow f(h(n)) = n$. □

3.5. Undecidability of the Halting Problem. The above corollary says that there is a single "universal" program which, given numbers e and $\vec{n}$, computes $\varphi_e^{(r)}(\vec{n})$ if it is defined. However we cannot decide in advance whether or not it will be defined. There is no program which, given e and $\vec{n}$, computes the total function

$$h(e, \vec{n}) \;=\; \begin{cases} 1 & \text{if } \varphi_e^{(r)}(\vec{n}) \text{ is defined,} \\ 0 & \text{if } \varphi_e^{(r)}(\vec{n}) \text{ is undefined.} \end{cases}$$

For suppose there were such a program. Then the function

$$\psi(\vec{n}) \;=\; \mu m\, (h(n_1, \vec{n}) = 0)$$

would be computable, say with fixed program number e_0, and therefore

$$\varphi_{e_0}^{(r)}(\vec{n}) \;=\; \begin{cases} 0 & \text{if } h(n_1, \vec{n}) = 0 \\ \text{undefined} & \text{if } h(n_1, \vec{n}) = 1 \end{cases}$$

But then fixing $n_1 = e_0$ gives:

$$\varphi_{e_0}^{(r)}(\vec{n}) \text{ defined} \iff h(e_0, \vec{n}) = 0 \iff \varphi_{e_0}^{(r)}(\vec{n}) \text{ undefined}$$

a contradiction. Hence the relation $R(e, \vec{n})$ which holds if and only if $\varphi_e^{(r)}(\vec{n})$ is defined, is not recursive. It is however Σ_1^0-definable.

There are numerous attempts to classify total computable functions according to the complexity of their termination proofs.

4. Recursive Definitions

4.1. Least Fixed Points of Recursive Definitions. By a *recursive definition* of a partial function φ of arity r from given partial functions $\psi_1, \ldots, \psi_m$ of fixed but unspecified arities, we mean a defining equation of the form

$$\varphi(n_1, \ldots, n_r) \;=\; t(\psi_1, \ldots, \psi_m, \varphi; n_1, \ldots, n_r)$$

where t is any compositional term built up from the numerical variables $\vec{n} = n_1, \ldots, n_r$ and the constant 0 by repeated applications of the successor and predecessor functions, the given functions $\psi_1, \ldots, \psi_m$, the function φ itself, and the "definition by cases" function :

$$\mathsf{dc}(x, y, u, v) \;=\; \begin{cases} u & \text{if } x, y \text{ are both defined and equal} \\ v & \text{if } x, y \text{ are both defined and unequal} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Our notion of recursive definition is essentially a reformulation of the Herbrand-Gödel-Kleene equation calculus; see Kleene [**15**].

There may be many partial functions φ satisfying such a recursive definition, but the one we wish to single out is the least defined one, i.e. the one whose defined values arise inevitably by *lazy evaluation* of the term t "from the outside in", making only those function calls which are absolutely necessary. This presupposes that each of the functions from which t is constructed already comes equipped with an evaluation strategy. In particular if a subterm $\mathsf{dc}(t_1, t_2, t_3, t_4)$ is called then it is to be evaluated according to the program construct:

$$x := t_1 \;\mathbf{;}\; y := t_2 \;\mathbf{;\ if}\; x := y \;\mathbf{then}\; t_3 \;\mathbf{else}\; t_4.$$

Some of the function calls demanded by the term t may be for further values of φ itself, and these must be evaluated by repeated unravellings of t (in other words by recursion).

This "least solution" φ will be referred to as *the function defined by that recursive definition* or its *least fixed point*. Its existence and its computability are guaranteed by Kleene's Recursion Theorem below.

4.2. The Principles of Finite Support and Monotonicity, and the Effective Index Property. Suppose we are given any fixed partial functions $\psi_1, \ldots, \psi_m$ and ψ, of the appropriate arities, and fixed inputs $\vec{n}$. If the term $t = t(\psi_1, \ldots, \psi_m, \psi; \vec{n})$ evaluates to a defined value k then the following principles are required to hold:

Finite Support Principle. Only finitely many values of $\psi_1, \ldots, \psi_m$ and ψ are used in that evaluation of t.

Monotonicity Principle. The same value k will be obtained no matter how the partial functions $\psi_1, \ldots, \psi_m$ and ψ are extended.

Note also that any such term t satisfies the

Effective Index Property. There is an elementary function f such that if $\psi_1, \ldots, \psi_m$ and ψ are computable partial functions with program numbers $e_1, \ldots, e_m$ and e respectively, then according to the lazy evaluation strategy just described,

$$t(\psi_1, \ldots, \psi_m, \psi; \vec{n})$$

defines a computable function of $\vec{n}$ with program number $f(e_1, \ldots, e_m, e)$.

The proof of the Effective Index Property is by induction over the build-up of the term t. The base case is where t is just one of the constants $0, 1$ or a variable n_j, in which case it defines either a constant function $\vec{n} \mapsto 0$ or $\vec{n} \mapsto 1$, or a projection function $\vec{n} \mapsto n_j$. Each of these is trivially computable with a fixed program number, and it is this program number we take as the value of $f(e_1, \ldots, e_m, e)$. Since in this case f is a constant function, it is clearly elementary. The induction step is where t is built up by applying one of the given functions: successor, predecessor, definition by cases or ψ (with or without a subscript) to previously constructed subterms $t_i(\psi_1, \ldots, \psi_m, \psi; \vec{n})$, $i = 1 \ldots l$, thus:

$$t \;=\; \psi(t_1, \ldots, t_l).$$

Inductively we can assume that for each $i = 1 \ldots l$, t_i defines a partial function of $\vec{n} = n_1, \ldots, n_r$ which is register machine computable by some program P_i with program number given by an already-constructed elementary function $f_i = f_i(e_1, \ldots, e_m, e)$. Therefore if ψ is computed by a program Q with program number e, we can put $P_1, \ldots, P_l$ and Q together to construct a new program obeying the evaluation strategy for t. Furthermore, by the remark on index-constructions near the beginning of Section 3, we will be able to compute its program number $f(e_1, \ldots, e_m, e)$ from the given numbers $f_1, \ldots, f_l$ and e, by some elementary function.

4.3. Recursion Theorem.

THEOREM (Kleene's Recursion Theorem). *For given partial functions $\psi_1, \ldots, \psi_m$, every recursive definition*

$$\varphi(\vec{n}) \;=\; t(\psi_1, \ldots, \psi_m, \varphi; \vec{n})$$

has a least fixed point, i.e. a least defined solution, φ. Moreover if $\psi_1, \ldots, \psi_m$ are computable, so is the least fixed point φ.

PROOF. Let $\psi_1, \ldots, \psi_m$ be fixed partial functions of the appropriate arities. Let Φ be the functional from partial functions of arity r to partial functions of arity r defined by lazy evaluation of the term t as described above:

$$\Phi(\psi)(\vec{n}) \;=\; t(\psi_1, \ldots, \psi_m, \psi; \vec{n}).$$

Let $\varphi_0, \varphi_1, \varphi_2, \ldots$ be the sequence of partial functions of arity r generated by Φ thus: φ_0 is the completely undefined function, and $\varphi_{i+1} = \Phi(\varphi_i)$ for each i. Then by induction on i, using the Monotonicity Principle above, we see that each φ_i is a subfunction of φ_{i+1}. That is, whenever $\varphi_i(\vec{n})$ is defined with a value k then $\varphi_{i+1}(\vec{n})$ is defined with that same value. Since their defined values are consistent with one another we can therefore construct the "union" φ of the φ_i's as follows:

$$\varphi(\vec{n}) = k \iff \exists i\, (\varphi_i(\vec{n}) = k).$$

(i) This φ is then the required least fixed point of the recursive definition.

To see that it is a fixed point, i.e. $\varphi = \Phi(\varphi)$, first suppose $\varphi(\vec{n})$ is defined with value k. Then by the definition of φ just given, there is an $i > 0$ such that $\varphi_i(\vec{n})$ is defined with value k. But $\varphi_i = \Phi(\varphi_{i-1})$ so $\Phi(\varphi_{i-1})(\vec{n})$ is defined with value k. Therefore by the Monotonicity Principle for Φ, since

φ_{i-1} is a subfunction of φ, $\Phi(\varphi)(\vec{n})$ is defined with value k. Hence φ is a subfunction of $\Phi(\varphi)$.

It remains to show the converse, that $\Phi(\varphi)$ is a subfunction of φ. So suppose $\Phi(\varphi)(\vec{n})$ is defined with value k. Then by the Finite Support Principle, only finitely many defined values of φ are called for in this evaluation. By the definition of φ there must be some i such that φ_i already supplies all of these required values, and so already at stage i we have $\Phi(\varphi_i)(\vec{n}) = \varphi_{i+1}(\vec{n})$ defined with value k. Since φ_{i+1} is a subfunction of φ it follows that $\varphi(\vec{n})$ is defined with value k. Hence $\Phi(\varphi)$ is a subfunction of φ.

To see that φ is the least such fixed point, suppose φ' is any fixed point of Φ. Then $\Phi(\varphi') = \varphi'$ so by the Monotonicity Principle, since φ_0 is a subfunction of φ' it follows that $\Phi(\varphi_0) = \varphi_1$ is a subfunction of $\Phi(\varphi') = \varphi'$. Then again by Monotonicity, $\Phi(\varphi_1) = \varphi_2$ is a subfunction of $\Phi(\varphi') = \varphi'$ etcetera so that for each i, φ_i is a subfunction of φ'. Since φ is the union of the φ_i's it follows that φ itself is a subfunction of φ'. Hence φ is the least fixed point of Φ.

(ii) Finally we have to show that φ is computable if the given functions $\psi_1, \ldots, \psi_m$ are. For this we need the Effective Index Property of the term t, which supplies an elementary function f such that if ψ is computable with program number e then $\Phi(\psi)$ is computable with program number $f(e) = f(e_1, \ldots, e_m, e)$. Thus if u is any fixed program number for the completely undefined function of arity r, $f(u)$ is a program number for $\varphi_1 = \Phi(\varphi_0)$, $f^2(u) = f(f(u))$ is a program number for $\varphi_2 = \Phi(\varphi_1)$, and in general $f^i(u)$ is a program number for φ_i. Therefore in the notation of the Normal Form Theorem,

$$\varphi_i(\vec{n}) \;=\; \varphi^{(r)}_{f^i(u)}(\vec{n})$$

and by the second corollary to the Normal Form Theorem, this is a computable function of i and $\vec{n}$, since $f^i(u)$ is a computable function of i definable (informally) say by a for-loop of the form "**for** $j = 1 \ldots i$ **do** f **od**". Therefore by the earlier equivalences, $\varphi_i(\vec{n})$ is a Σ^0_1-definable function of i and $\vec{n}$, and hence so is φ itself because

$$\varphi(\vec{n}) = m \iff \exists i \, (\, \varphi_i(\vec{n}) = m \,) \, .$$

So φ is computable and this completes the proof. □

NOTE. The above proof works equally well if φ is a vector-valued function. In other words if, instead of defining a single partial function φ, the recursive definition in fact defines a finite list $\vec{\varphi}$ of such functions *simultaneously*. For example, the individual components of the machine state of any register machine at step s are clearly defined by a simultaneous recursive definition, from zero and successor.

4.4. Recursive Programs and Partial Recursive Functions. A *recursive program* is a finite sequence of possibly simultaneous recursive definitions:

$$\begin{aligned}
\vec{\varphi}_0(n_1, \ldots, n_{r_0}) &= t_0(\vec{\varphi}_0; n_1, \ldots, n_{r_0}) \\
\vec{\varphi}_1(n_1, \ldots, n_{r_1}) &= t_1(\vec{\varphi}_0, \vec{\varphi}_1; n_1, \ldots, n_{r_1}) \\
\vec{\varphi}_2(n_1, \ldots, n_{r_2}) &= t_2(\vec{\varphi}_0, \vec{\varphi}_1, \vec{\varphi}_2; n_1, \ldots, n_{r_2})
\end{aligned}$$

$$\vdots$$

$$\vec{\varphi}_k(n_1, \ldots, n_{r_k}) = t_k(\vec{\varphi}_0, \ldots, \vec{\varphi}_{k-1}, \vec{\varphi}_k; n_1, \ldots, n_{r_k}).$$

A partial function is said to be *partial recursive* if it is one of the functions defined by some recursive program as above. A partial recursive function which happens to be totally defined is called simply a *recursive function*.

THEOREM. *A function is partial recursive if and only if it is computable.*

PROOF. The Recursion Theorem tells us immediately that every partial recursive function is computable. For the converse we use the equivalence of computability with μ-recursiveness already established in Section 3. Thus we need only show how to translate any μ-recursive definition into a recursive program:

The constant 0 function is defined by the recursive program

$$\varphi(\vec{n}) \;=\; 0$$

and similarly for the constant 1 function.

The addition function $\varphi(m, n) = m + n$ is defined by the recursive program

$$\varphi(m, n) \;=\; \mathsf{dc}(n, 0, m, \varphi(m, n \dot{-} 1) + 1)$$

and the subtraction function $\varphi(m, n) = m \dot{-} n$ is defined similarly but with the successor function $+1$ replaced by the predecessor $\dot{-}1$. Multiplication is defined recursively from addition in much the same way. Note that in each case the right hand side of the recursive definition is an allowed term.

The composition scheme is a recursive definition as it stands.

Finally, given a recursive program defining ψ, if we add to it the recursive definition:

$$\varphi(\vec{n}, m) \;=\; \mathsf{dc}(\, \psi(\vec{n}, m), 0, m, \varphi(\vec{n}, m + 1)\,)$$

followed by

$$\varphi'(\vec{n}) \;=\; \varphi(\vec{n}, 0)$$

then the computation of $\varphi'(\vec{n})$ proceeds as follows:

$$\begin{aligned}
\varphi'(\vec{n}) &= \varphi(\vec{n}, 0) \\
&= \varphi(\vec{n}, 1) \quad \text{if } \psi(\vec{n}, 0) \neq 0 \\
&= \varphi(\vec{n}, 2) \quad \text{if } \psi(\vec{n}, 1) \neq 0 \\
&\;\;\vdots \\
&= \varphi(\vec{n}, m) \quad \text{if } \psi(\vec{n}, m-1) \neq 0 \\
&= m \qquad\quad \text{if } \psi(\vec{n}, m) = 0
\end{aligned}$$

Thus the recursive program for φ' defines unbounded minimization:

$$\varphi'(\vec{n}) \;=\; \mu m\, (\psi(\vec{n}, m) = 0).$$

This completes the proof. □

CHAPTER 4

Gödel's Theorems

1. Gödel Numbers

1.1. Coding Terms and Formulas. We use the elementary sequence-coding and decoding machinery developed earlier. Let $\mathcal{L}$ be a countable first order language. Assume that we have injectively assigned to every n-ary relation symbol R a *symbol number* $\mathsf{SN}(R)$ of the form $\langle 1, n, i\rangle$ and to every n-ary function symbol f a symbol number $\mathsf{SN}(f)$ of the form $\langle 2, n, j\rangle$. Call $\mathcal{L}$ *elementarily presented*, if the set $\mathsf{Symb}_{\mathcal{L}}$ of all these symbol numbers is elementary. In what follows we shall always assume that the languages $\mathcal{L}$ considered are elementarily presented. In particular this applies to every language with finitely many relation and function symbols.

Assign numbers to the logical symbols by $\mathsf{SN}(\wedge) := \langle 3, 1\rangle$, $\mathsf{SN}(\to) := \langle 3, 2\rangle$ und $\mathsf{SN}(\forall) := \langle 3, 3\rangle$, and to the i-th variable assign the symbol number $\langle 0, i\rangle$.

For every $\mathcal{L}$-term t we define recursively its Gödel number $\ulcorner t \urcorner$ by

$$\begin{aligned}
\ulcorner x \urcorner &:= \langle \mathsf{SN}(x)\rangle,\\
\ulcorner c \urcorner &:= \langle \mathsf{SN}(c)\rangle,\\
\ulcorner f t_1 \dots t_n \urcorner &:= \langle \mathsf{SN}(f), \ulcorner t_1 \urcorner, \dots, \ulcorner t_n \urcorner\rangle.
\end{aligned}$$

Similarly we recursively define for every $\mathcal{L}$-formula A its Gödel number $\ulcorner A \urcorner$ by

$$\begin{aligned}
\ulcorner R t_1 \dots t_n \urcorner &:= \langle \mathsf{SN}(R), \ulcorner t_1 \urcorner, \dots, \ulcorner t_n \urcorner\rangle,\\
\ulcorner A \wedge B \urcorner &:= \langle \mathsf{SN}(\wedge), \ulcorner A \urcorner, \ulcorner B \urcorner\rangle,\\
\ulcorner A \to B \urcorner &:= \langle \mathsf{SN}(\to), \ulcorner A \urcorner, \ulcorner B \urcorner\rangle,\\
\ulcorner \forall x\, A \urcorner &:= \langle \mathsf{SN}(\forall), \ulcorner x \urcorner, \ulcorner A \urcorner\rangle.
\end{aligned}$$

Let $\mathsf{Var} := \{\, \langle\langle 0, i\rangle\rangle \mid i \in \mathbb{N} \,\}$. Var clearly is elementary, and we have $a \in \mathsf{Var}$ if and only if $a = \ulcorner x \urcorner$ for a variable x. We define $\mathsf{Ter} \subseteq \mathbb{N}$ as follows, by course-of-values recursion.

$$\begin{aligned}
&a \in \mathsf{Ter} :\leftrightarrow\\
&\quad a \in \mathsf{Var} \vee\\
&\quad ((a)_0 \in \mathsf{Symb}_{\mathcal{L}} \wedge (a)_{0,0} = 2 \wedge \mathsf{lh}(a) = (a)_{0,1} + 1 \wedge \forall i_{0<i<\mathsf{lh}(a)}\, (a)_i \in \mathsf{Ter}).
\end{aligned}$$

Ter is elementary, and it is easily seen that $a \in \mathsf{Ter}$ if and only if $a = \ulcorner t \urcorner$ for some term t. Similarly $\mathsf{For} \subseteq \mathbb{N}$ is defined by

$$\begin{aligned}
&a \in \mathsf{For} :\leftrightarrow\\
&\quad ((a)_0 \in \mathsf{Symb}_{\mathcal{L}} \wedge (a)_{0,0} = 1 \wedge \mathsf{lh}(a) = (a)_{0,1} + 1 \wedge \forall i_{0<i<\mathsf{lh}(a)}\, (a)_i \in \mathsf{Ter}) \vee\\
&\quad (a = \langle \mathsf{SN}(\wedge), (a)_1, (a)_2\rangle \wedge (a)_1 \in \mathsf{For} \wedge (a)_2 \in \mathsf{For}) \vee
\end{aligned}$$

$$(a = \langle \mathsf{SN}(\to), (a)_1, (a)_2\rangle \wedge (a)_1 \in \mathsf{For} \wedge (a)_2 \in \mathsf{For}) \vee$$
$$(a = \langle \mathsf{SN}(\forall), (a)_1, (a)_2\rangle \wedge (a)_1 \in \mathsf{Var} \wedge (a)_2 \in \mathsf{For}).$$

Again For is elementary, and we have $a \in \mathsf{For}$ if and only if $a = \ulcorner A \urcorner$ for some formula A. For a set S of formulas let $\ulcorner S \urcorner := \{\, \ulcorner A \urcorner \mid A \in S \,\}$.

We could continue in this way and define Gödel numberings of various other syntactical notions, but we are mainly concerned that the reader believes that it *can* be done rather than sees all the (gory) details. In particular there are elementary functions msm and sub with the following properties:

- $\mathsf{msm}(\ulcorner \Gamma \urcorner, \ulcorner \Delta \urcorner)$ codes the result of deleting from the multiset of formulas Γ all the formulas which lie in Δ.
- $\mathsf{sub}(\ulcorner t \urcorner, \ulcorner \vartheta \urcorner) = \ulcorner t\vartheta \urcorner$, and $\mathsf{sub}(\ulcorner A \urcorner, \ulcorner \vartheta \urcorner) = \ulcorner A\vartheta \urcorner$, where ϑ is a substitution (i.e. a finite assignment of terms to variables) and $t\vartheta$ and $A\vartheta$ are the results of substituting those terms for those variables in t or A respectively.

1.2. Sequents. In our previous exposition of natural deduction one can find the assumptions free at a given node by inspecting the upper part of the proof tree. An alternative is to write the free assumptions next to each node, in the form of a multiset.

By a *sequent* $\Gamma \Rightarrow A$ we mean a pair consisting of a multiset $\Gamma = \{\{A_1, \dots, A_n\}\}$ of formulas and a single formula A. We define $\vdash_m \Gamma \Rightarrow A$ inductively by the following rules. An assumption can be introduced by

$$\Gamma \Rightarrow A \quad \text{if } A \text{ in } \Gamma.$$

For conjunction $\wedge$ we have an introduction rule $\wedge$I and two elimination rules $\wedge \mathrm{E_l}$ und $\wedge \mathrm{E_r}$.

$$\frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A \wedge B}\wedge \mathrm{I} \qquad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow A}\wedge \mathrm{E_r} \qquad \frac{\Gamma \Rightarrow A \wedge B}{\Gamma \Rightarrow B}\wedge \mathrm{E_l}$$

Here Γ, Δ denotes multiset union. For implication $\to$ we have an introduction rule $\to$I (not mentioning an assumption variable u) and an elimination rule $\to$E.

$$\frac{\Gamma \Rightarrow B}{\Delta \Rightarrow A \to B}\to \mathrm{I} \qquad \frac{\Gamma \Rightarrow A \to B \quad \Delta \Rightarrow A}{\Gamma, \Delta \Rightarrow B}\to \mathrm{E}$$

In $\to$I the multiset Δ is obtained from Γ by cancelling some occurrences of A. For the universal quantifier $\forall$ we have an introduction rule $\forall$I and an elimination rule $\forall$E (formulated without the term t to be substituted as additional premise)

$$\frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \forall x A}\forall \mathrm{I} \qquad \frac{\Gamma \Rightarrow \forall x A}{\Gamma \Rightarrow A[x := t]}\forall \mathrm{E}$$

In $\forall$I the variable condition needs to hold: for all B in Γ we must have $x \notin \mathsf{FV}(B)$.

LEMMA. (a) *If* $\vdash_m \{\{A_1, \dots, A_n\}\} \Rightarrow A$, *then for all (not necessarily distinct)* $u_1, \dots, u_n$ *such that* $u_i = u_j \to A_i = A_j$ *we can find a derivation term* $M^A[u_1^{A_1}, \dots, u_n^{A_n}]$.

(b) *For every derivation term* $M^A[u_1^{A_1}, \dots, u_n^{A_n}]$ *one can find multiplicities* $k_1, \dots, k_n \geq 0$ *such that* $\vdash_m \{\{A_1^{k_1}, \dots, A_n^{k_n}\}\} \Rightarrow A$; *here* A^k *means a* k*-fold occurrence of* A.

PROOF. (a). Assume $\vdash_m \{\{A_1,\dots,A_n\}\} \Rightarrow A$. We use induction on $\vdash_m$. **Case** assumption. Let $A = A_i$. Take $M = u_i$.

Case $\to$I. We may assume

$$\frac{\{\{A_1,\dots,A_n,A,\dots,A\}\} \Rightarrow B}{\{\{A_1,\dots,A_n\}\} \Rightarrow A \to B}\to\text{I}.$$

Let $u_1,\dots,u_n$ be given such that $u_i = u_j \to A_i = A_j$. Pick a new assumption variable u. By IH there exists an M^B such that $\mathsf{FA}(M^B) \subseteq \{u_1^{A_1},\dots,u_n^{A_n},u^A\}$. Then $(\lambda u^A\, M^B)^{A\to B}$ is a derivation term with free assumptions among $u_1^{A_1},\dots,u_n^{A_n}$.

Case $\to$E. Assume we have derivations of

$$\{\{A_1,\dots,A_n\}\} \Rightarrow A \to B \quad\text{and}\quad \{\{A_{n+1},\dots,A_{n+m}\}\} \Rightarrow A.$$

Let $u_1,\dots,u_{n+m}$ be given such that $u_i = u_j \to A_i = A_j$. By IH we have derivation terms

$$M^{A\to B}[u_1^{A_1},\dots,u_n^{A_n}] \quad\text{and}\quad N^A[u_{n+1}^{A_{n+1}},\dots,u_{n+m}^{A_{n+m}}].$$

But then also

$$(MN)^B[u_1^{A_1},\dots,u_{n+m}^{A_{n+m}}]$$

is a derivation term.

The other cases are treated similarly.

(b). Let a derivation term $M^A[u_1^{A_1},\dots,u_n^{A_n}]$ be given. We use induction on M. **Case** u^A. Then $\vdash_m \{\{A\}\} \Rightarrow A$.

Case $\to$I, so $(\lambda u^A M^B)^{A\to B}$. Let $\mathsf{FA}(M^B) \subseteq \{u_1^{A_1},\dots,u_n^{A_n},u^A\}$ with $u_1,\dots,u_n,u$ distinct. By IH we have

$$\vdash_m \{\{A_1^{k_1},\dots,A_n^{k_n},A^k\}\} \Rightarrow B.$$

Using the rule $\to$I we obtain

$$\vdash_m \{\{A_1^{k_1},\dots,A_n^{k_n}\}\} \Rightarrow A \to B.$$

Case $\to$E. We are given $(M^{A\to B}N^A)^B[u_1^{A_1},\dots,u_n^{A_n}]$. By IH we have

$$\vdash_m \{\{A_1^{k_1},\dots,A_n^{k_n}\}\} \Rightarrow A \to B \quad\text{and}\quad \vdash_m \{\{A_1^{l_1},\dots,A_n^{l_n}\}\} \Rightarrow A.$$

Using the rule $\to$E we obtain

$$\{\{A_1^{k_1+l_1},\dots,A_n^{k_n+l_n}\}\} \Rightarrow B.$$

The other cases are treated similarly. □

1.3. Coding Derivations. We can now define the set of Gödel numbers of formal proofs (in the above formulation of minimal logic), as follows.

$$\begin{aligned}
&\mathsf{Deriv}(d) :\leftrightarrow \forall i{<}\mathsf{lh}(d).\\
&(\forall m{<}\mathsf{lh}((d)_{i,0})\,\mathsf{For}((d)_{i,0,m}) \wedge \exists n{<}\mathsf{lh}((d)_{i,0})\,((d)_{i,1} = (d)_{i,0,n})) && (\text{A})\\
&\vee (\exists j,k{<}i.(d)_{i,1} = \langle \mathsf{SN}(\wedge),(d)_{j,1},(d)_{k,1}\rangle \wedge (d)_{i,0} =_m (d)_{j,0} * (d)_{k,0}) && (\wedge\text{I})\\
&\vee (\exists j{<}i.(d)_{j,1} = \langle \mathsf{SN}(\wedge),(d)_{i,1},(d)_{j,1,2}\rangle \wedge (d)_{i,0} =_m (d)_{j,0}) && (\wedge\text{E}_\text{r})\\
&\vee (\exists j{<}i.(d)_{j,1} = \langle \mathsf{SN}(\wedge),(d)_{j,1,1},(d)_{i,1}\rangle \wedge (d)_{i,0} =_m (d)_{j,0}) && (\wedge\text{E}_\text{l})\\
&\vee (\exists j{<}i.(d)_{i,1} = \langle \mathsf{SN}(\to),(d)_{i,1,1},(d)_{j,1}\rangle \wedge \mathsf{For}((d)_{i,1,1}) && (\to\text{I})\\
&\quad \wedge \mathsf{msm}((d)_{i,0},(d)_{j,0}) = 0\\
&\quad \wedge \forall n{<}\mathsf{lh}(\mathsf{msm}((d)_{j,0},(d)_{i,0}))\,((\mathsf{msm}((d)_{j,0},(d)_{i,0}))_n = (d)_{i,1,1}))
\end{aligned}$$

$$\vee\,(\exists j,k{<}i.(d)_{j,1} = \langle \mathsf{SN}(\to),(d)_{k,1},(d)_{i,1}\rangle \wedge (d)_{i,0} =_m (d)_{j,0} * (d)_{k,0}) \quad (\to\mathrm{E})$$
$$\vee\,(\exists j{<}i.(d)_{i,1} = \langle \mathsf{SN}(\forall),(d)_{i,1,1},(d)_{j,1}\rangle \wedge \mathsf{Var}((d)_{i,1,1}) \quad (\forall\mathrm{I})$$
$$\wedge\,(d)_{i,0} =_m (d)_{j,0} \wedge \forall n{<}\mathsf{lh}((d)_{i,0})\,\neg\mathsf{FV}((d)_{i,1,1},(d)_{i,0,n}))$$
$$\vee\,(\exists j{<}i.(d)_{j,1,0} = \mathsf{SN}(\forall) \wedge (d)_{i,0} =_m (d)_{j,0} \quad (\forall\mathrm{E})$$
$$\wedge\,((d)_{i,1} = (d)_{j,1,2} \vee \exists n{<}(d)_{i,1}.\mathsf{Ter}(n)$$
$$\wedge\,(d)_{i,1} = \mathsf{sub}((d)_{j,1,2},\langle\langle(d)_{j,1,1},n\rangle\rangle))).$$

Note that (1) this clearly defines an elementary set, and (2) if one carefully reads the definition, then it becomes clear that d is in Deriv iff d codes a sequence of pairs (sequents) $\Gamma_i \Rightarrow A_i$ with Γ_i a multiset, such that this sequence constitutes a derivation in minimal logic, i.e. each sequent is either an axiom or else follows from previous sequents by a rule. Thus

LEMMA.
(a) $\mathsf{Deriv}(d)$ *if and only if* d *is the Gödel number of a derivation.*
(b) Deriv *is elementary.*

1.4. Axiomatizable Theories. A set S of formulas is called *recursive* (*elementary*, Σ^0_1*-definable*), if $\ulcorner S\urcorner := \{\ulcorner A\urcorner \mid A \in S\}$ is recursive (elementary, Σ^0_1-definable). Clearly the sets $\mathsf{Stabax}_{\mathcal{L}}$ of stability axioms and $\mathsf{Eq}_{\mathcal{L}}$ of $\mathcal{L}$-equality axioms are elementary.

Now let $\mathcal{L}$ be an elementarily presented language with $=$ in $\mathcal{L}$. A theory T with $L(T) \subseteq \mathcal{L}$ is called *recursively (elementarily) axiomatizable*, if there is a recursive (elementary) set S of closed $\mathcal{L}$-formulas such that $T = \{A \in \overline{\mathcal{L}} \mid S \cup \mathsf{Eq}_{\mathcal{L}} \vdash_c A\}$.

THEOREM. *For theories* T *with* $L(T) \subseteq \mathcal{L}$ *the following are equivalent.*
(a) T *is recursively axiomatizable.*
(b) T *is elementarily axiomatizable.*
(c) T *is* Σ^0_1*-definable.*

PROOF. (c) $\Rightarrow$ (b). Let $\ulcorner T\urcorner$ be Σ^0_1-definable. Then by Section 2.5 of Chapter 3 there exists an $f \in \mathcal{E}$ such that $\ulcorner T\urcorner = \mathsf{ran}(f)$, and by the argument there we can assume $f(n) \leq n$ for all n. Let $f(n) = \ulcorner A_n\urcorner$. We define an elementary function g with the property $g(n) = \ulcorner A_0 \wedge \cdots \wedge A_n\urcorner$ by

$$\begin{aligned} g(0) &:= f(0),\\ g(n+1) &:= g(n) \mathbin{\dot\wedge} f(n+1),\end{aligned}$$

where $a \mathbin{\dot\wedge} b := \langle \mathsf{SN}(\wedge),a,b\rangle$. Clearly g can be bounded in $\mathcal{E}$. For $S := \{A_0 \wedge \cdots \wedge A_n \mid n \in \mathbb{N}\}$ we have $\ulcorner S\urcorner = \mathsf{ran}(g)$, and this set is elementary because of $a \in \mathsf{ran}(g) \leftrightarrow \exists n{<}a\,(a = g(n))$. T is elementarily axiomatizable, since $T = \{A \in \overline{\mathcal{L}} \mid S \cup \mathsf{Eq}_{\mathcal{L}} \vdash_c A\}$.

(b) $\Rightarrow$ (a) is clear.

(a) $\Rightarrow$ (c). Let T be axiomatized by S with $\ulcorner S\urcorner$ recursive. Then

$$a \in \ulcorner T\urcorner \leftrightarrow \exists d\exists c{<}d.\mathsf{Deriv}(d) \wedge (d)_{\mathsf{lh}(d)\dot-1} = \langle c,a\rangle \wedge \forall i{<}\mathsf{lh}(c)\,((c)_i \in \ulcorner\mathsf{Stabax}\urcorner \cup \ulcorner\mathsf{Eq}\urcorner \cup \ulcorner S\urcorner).$$

Hence $\ulcorner T\urcorner$ is Σ^0_1-definable. □

A theory T in our elementarily presented language $\mathcal{L}$ is called *axiomatized*, if it is given by a Σ^0_1-definable axiom system Ax_T. By the theorem just proved we can even assume that Ax_T is elementary. For such axiomatized theories we define $\mathsf{Prf}_T \subseteq \mathbb{N} \times \mathbb{N}$ by

$$\mathsf{Prf}_T(d,a) :\leftrightarrow \mathsf{Deriv}(d) \wedge \exists c{<}d.(d)_{\mathsf{lh}(d)\dot{-}1} = \langle c, a\rangle \wedge \\ \forall i{<}\mathsf{lh}(c)\,((c)_i \in \ulcorner \mathsf{Stabax} \urcorner \cup \ulcorner \mathsf{Eq} \urcorner \cup \ulcorner \mathsf{Ax}_T \urcorner).$$

Clearly Prf_T is elementary and $\mathsf{Prf}_T(d,a)$ if and only if d is a derivation of a sequent $\Gamma \Rightarrow A$ with Γ composed from stability axioms, equality axioms and formulas from Ax_T, and $a = \ulcorner A \urcorner$.

A theory T is called *consistent*, if there is a closed formula A such that $A \notin T$; otherwise T is called *inconsistent*.

Corollary. *Every axiomatized complete theory T is recursive.*

Proof. If T is inconsistent, then $\ulcorner T \urcorner$ is recursive. If not, then from the completeness of T we obtain

$$a \in \mathbb{N} \setminus \ulcorner T \urcorner \leftrightarrow a \notin \mathsf{For} \vee \exists b{<}a\, \mathsf{FV}(b,a) \vee \dot{\neg} a \in \ulcorner T \urcorner,$$

where $\dot{\neg} a := a \mathrel{\dot{\to}} \ulcorner \bot \urcorner$ and $a \mathrel{\dot{\to}} b := \langle \mathsf{SN}(\to), a, b\rangle$. Hence with $\ulcorner T \urcorner$ also $\mathbb{N} \setminus \ulcorner T \urcorner$ is Σ^0_1-definable and therefore $\ulcorner T \urcorner$ is recursive. □

2. Undefinability of the Notion of Truth

Recall the convention in 1.2 of Chapter 1: once a formula has been introduced as $A(x)$, i.e., A with a designated variable x, we write $A(t)$ for $A[x := t]$, and similarly with more variables.

2.1. Definable Relations. Let $\mathcal{M}$ be an $\mathcal{L}$-structure. A relation $R \subseteq |\mathcal{M}|^n$ is called *definable* in $\mathcal{M}$ if there is an $\mathcal{L}$-formula $A(x_1, \dots, x_n)$ with only the free variables shown such that

$$R = \{\, (a_1, \dots, a_n) \in |\mathcal{M}|^n \mid \mathcal{M} \models A[a_1, \dots, a_n] \,\}.$$

We assume in this section that $|\mathcal{M}| = \mathbb{N}$, 0 is a constant in $\mathcal{L}$ and S is a unary function symbol in $\mathcal{L}$ with $0^{\mathcal{M}} = 0$ and $\mathsf{S}^{\mathcal{M}}(a) = a{+}1$. Then for every $a \in \mathbb{N}$ we can define the *numeral* $\underline{a} \in \mathsf{Ter}_{\mathcal{L}}$ by $\underline{0} := 0$ and $\underline{a+1} := \mathsf{S}(\underline{a})$. Observe that in this case the definability of $R \subseteq \mathbb{N}^n$ by $A(x_1, \dots, x_n)$ is equivalent to

$$R = \{\, (a_1, \dots, a_n) \in \mathbb{N}^n \mid \mathcal{M} \models A(\underline{a_1}, \dots, \underline{a_n}) \,\}.$$

Furthermore let $\mathcal{L}$ be an elementarily presented language. We shall always assume in this section that every elementary relation is definable in $\mathcal{M}$. A set S of formulas is called *definable* in $\mathcal{M}$, if $\ulcorner S \urcorner := \{\, \ulcorner A \urcorner \mid A \in S \,\}$ is definable in $\mathcal{M}$.

We shall show that already from these assumptions it follows that the notion of truth for $\mathcal{M}$, more precisely the set $\mathsf{Th}(\mathcal{M})$ of all closed formulas valid in $\mathcal{M}$, is undefinable in $\mathcal{M}$. From this it will follow in turn that the notion of truth is in fact undecidable, for otherwise the set $\mathsf{Th}(\mathcal{M})$ would be recursive (by Church's Thesis), hence Σ^0_1-definable, and hence definable, because we have assumed already that all elementary relations are definable in $\mathcal{M}$.

2.2. Fixed Points. For the proof we shall need the following lemma, which will be generalized in the next section.

LEMMA (Semantical Fixed Point Lemma). *If every elementary relation is definable in $\mathcal{M}$, then for every $\mathcal{L}$-formula $B(z)$ with only z free we can find a closed $\mathcal{L}$-formula A such that*

$$\mathcal{M} \models A \quad \textit{if and only if} \quad \mathcal{M} \models B[\ulcorner A \urcorner].$$

PROOF. We define an elementary function s by

$$s(b,k) := \mathsf{sub}(b, \langle\langle \ulcorner z \urcorner, \ulcorner \underline{k} \urcorner\rangle\rangle).$$

Here z is a specially given variable determined by $B(z)$, say $*_0$. Then for every formula $C(z)$ we have

$$s(\ulcorner C \urcorner, k) = \mathsf{sub}(\ulcorner C \urcorner, \langle\langle \ulcorner z \urcorner, \ulcorner \underline{k} \urcorner\rangle\rangle) = \ulcorner C(\underline{k}) \urcorner,$$

hence in particular

$$s(\ulcorner C \urcorner, \ulcorner C \urcorner) = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner.$$

By assumption the graph G_s of s is definable in $\mathcal{M}$, by $A_s(x_1, x_2, x_3)$ say. Let

$$C(z) := \exists x.B(x) \wedge A_s(z,z,x),$$
$$A := C(\underline{\ulcorner C \urcorner}),$$

so

$$A = \exists x.B(x) \wedge A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x).$$

Hence $\mathcal{M} \models A$ if and only if $\exists a \in \mathbb{N}.\mathcal{M} \models B[a]$ and $a = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner$, so if and only if $\mathcal{M} \models B[\ulcorner A \urcorner]$. □

2.3. Undefinability. We can now prove the undefinability of truth.

THEOREM (Tarski's Undefinability Theorem). *Assume that every elementary relation is definable in $\mathcal{M}$. Then $\mathsf{Th}(\mathcal{M})$ is undefinable in $\mathcal{M}$, hence in particular not Σ^0_1-definable.*

PROOF. Assume that $\ulcorner \mathsf{Th}(\mathcal{M}) \urcorner$ is definable by $B_W(z)$. Then for all closed formulas A

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models B_W[\ulcorner A \urcorner].$$

Now consider the formula $\neg B_W(z)$ and choose by the Fixed Point Lemma a closed $\mathcal{L}$-formula A such that

$$\mathcal{M} \models A \quad \text{if and only if} \quad \mathcal{M} \models \neg B_W[\ulcorner A \urcorner].$$

This contradicts the equivalence above.

We already have noticed that all Σ^0_1-definable relations are definable in $\mathcal{M}$. Hence it follows that $\ulcorner \mathsf{Th}(\mathcal{M}) \urcorner$ cannot be Σ^0_1-definable. □

3. The Notion of Truth in Formal Theories

We now want to generalize the arguments of the previous section. There we have made essential use of the notion of truth in a structure $\mathcal{M}$, i.e. of the relation $\mathcal{M} \models A$. The set of all closed formulas A such that $\mathcal{M} \models A$ has been called the theory of $\mathcal{M}$, denoted $\mathsf{Th}(\mathcal{M})$.

Now instead of $\mathsf{Th}(\mathcal{M})$ we shall start more generally from an arbitrary theory T. We shall deal with the question as to whether in T there is a *notion of truth* (in the form of a *truth formula* $B(z)$), such that $B(z)$ "means" that z is "true".

What shall this mean? We have to explain all the notions used without referring to semantical concepts at all.

- z ranges over closed formulas (or sentences) A, or more precisely over their Gödel numbers $\ulcorner A \urcorner$.
- A "true" is to be replaced by $T \vdash A$.
- C "equivalent" to D is to be replaced by $T \vdash C \leftrightarrow D$.

We want to study the question as to whether it is possible that a truth formula $B(z)$ exists, such that for all sentences A we have $T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner})$. The result will be that this is impossible, under rather weak assumptions on the theory T.

3.1. Representable Relations. Technically, the issue will be to replace the notion of definability by the notion of "representability" within a formal theory.

Let $\mathcal{L}$ again be an elementarily presented language with $0, \mathsf{S}, =$ in $\mathcal{L}$ and T be a theory containing the equality axioms $\mathsf{Eq}_{\mathcal{L}}$.

Definition. A relation $R \subseteq \mathbb{N}^n$ is *representable* in T if there is a formula $A(x_1, \ldots, x_n)$ such that

$$\begin{aligned} T &\vdash A(\underline{a_1}, \ldots, \underline{a_n}), && \text{if } (a_1, \ldots, a_n) \in R, \\ T &\vdash \neg A(\underline{a_1}, \ldots, \underline{a_n}), && \text{if } (a_1, \ldots, a_n) \notin R. \end{aligned}$$

A function $f\colon \mathbb{N}^n \to \mathbb{N}$ is called *representable* in T if there is a formula $A(x_1, \ldots, x_n, y)$ representing the graph $G_f \subseteq \mathbb{N}^{n+1}$ of f, i.e., such that

$$T \vdash A(\underline{a_1}, \ldots, \underline{a_n}, \underline{f(a_1, \ldots, a_n)}), \tag{16}$$

$$T \vdash \neg A(\underline{a_1}, \ldots, \underline{a_n}, \underline{c}), \qquad \text{if } c \neq f(a_1, \ldots, a_n) \tag{17}$$

and such that in addition

$$T \vdash A(\underline{a_1}, \ldots, \underline{a_n}, y) \to A(\underline{a_1}, \ldots, \underline{a_n}, z) \to y = z \text{ for all } a_1, \ldots, a_n \in \mathbb{N}. \tag{18}$$

Notice that in case $T \vdash \underline{b} \neq \underline{c}$ for $b < c$ the condition (17) follows from (16) and (18).

Lemma. *If the characteristic function c_R of a relation $R \subseteq \mathbb{N}^n$ is representable in T, then so is the relation R itself.*

Proof. For simplicity assume $n = 1$. Let $A(x, y)$ be a formula representing c_R. We show that $A(x, \underline{1})$ represents the relation R. So assume $a \in R$. Then $c_R(a) = 1$, hence $(a, 1) \in G_{c_R}$, hence $T \vdash A(\underline{a}, \underline{1})$. Conversely, assume $a \notin R$. Then $c_R(a) = 0$, hence $(a, 1) \notin G_{c_R}$, hence $T \vdash \neg A(\underline{a}, \underline{1})$. $\square$

3.2. Fixed Points. We can now prove a generalized (syntactical) version of the Fixed Point Lemma above.

LEMMA (Fixed Point Lemma). *Assume that all elementary functions are representable in T. Then for every formula $B(z)$ with only z free we can find a closed formula A such that*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

PROOF. We start as in the proof of the Semantical Fixed Point Lemma. Let $A_s(x_1, x_2, x_3)$ be a formula which represents the elementary function $s(b, k) := \mathsf{sub}(b, \langle\langle \ulcorner z \urcorner, \ulcorner \underline{k} \urcorner\rangle\rangle)$. Let

$$C(z) := \exists x.B(x) \wedge A_s(z, z, x),$$
$$A := C(\underline{\ulcorner C \urcorner}),$$

i.e.

$$A = \exists x.B(x) \wedge A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x).$$

Because of $s(\ulcorner C \urcorner, \ulcorner C \urcorner) = \ulcorner C(\underline{\ulcorner C \urcorner}) \urcorner = \ulcorner A \urcorner$ we can prove in T

$$A_s(\underline{\ulcorner C \urcorner}, \underline{\ulcorner C \urcorner}, x) \leftrightarrow x = \underline{\ulcorner A \urcorner},$$

hence by definition of A also

$$A \leftrightarrow \exists x.B(x) \wedge x = \underline{\ulcorner A \urcorner}$$

and hence

$$A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

□

Notice that for $T = \mathsf{Th}(\mathcal{M})$ we obtain the above (semantical) Fixed Point Lemma as a special case.

3.3. Undefinability. Using the Fixed Point Lemma above, we can generalize the undefinability result as well.

THEOREM (Undefinability of the Notion of Truth). *Let T be a consistent theory such that all elementary functions are representable in T. Then there cannot exist a formula $B(z)$ with only z free defining the notion of truth, i.e. such that for all closed formulas A*

$$T \vdash A \leftrightarrow B(\underline{\ulcorner A \urcorner}).$$

PROOF. Assume we would have such a $B(z)$. Consider the formula $\neg B(z)$ and choose by the Fixed Point Lemma a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

For this A we have $T \vdash A \leftrightarrow \neg A$, contradicting the consistency of T. □

For $T = \mathsf{Th}(\mathcal{M})$ Tarski's Undefinability Theorem is a special case.

4. Undecidability and Incompleteness

In this section we consider a consistent formal theory T with the property that all recursive functions are representable in T. This is a very weak assumption, as we shall show in the next section: it is always satisfied if the theory allows to develop a certain minimum of arithmetic.

We shall show that such a theory necessarily is undecidable. Moreover we shall prove Gödel's First Incompleteness Theorem, which says that every axiomatized such theory must be incomplete. We will also prove a sharpened form of this theorem due to Rosser, which explicitely provides a closed formula A such that neither A nor $\neg A$ is provable in the theory T.

4.1. Undecidability; Gödel's First Incompleteness Theorem. Let again $\mathcal{L}$ be an elementarily presented language with $0, \mathsf{S}, =$ in $\mathcal{L}$, and T be a theory containing the equality axioms $\mathsf{Eq}_{\mathcal{L}}$.

THEOREM. *Assume that T is a consistent theory such that all recursive functions are representable in T. Then T is not recursive.*

PROOF. Assume that T is recursive. By assumption there exists a formula $B(z)$ in representing $\ulcorner T \urcorner$ in T. Choose by the Fixed Point Lemma in 3.2 a closed formula A such that

$$T \vdash A \leftrightarrow \neg B(\underline{\ulcorner A \urcorner}).$$

We shall prove $(*)$ $T \nvdash A$ and $(**)$ $T \vdash A$; this is the desired contradiction.

Ad $(*)$. Assume $T \vdash A$. Then $A \in T$, hence $\ulcorner A \urcorner \in \ulcorner T \urcorner$, hence $T \vdash B(\underline{\ulcorner A \urcorner})$ (because $B(z)$ represents in T the set $\ulcorner T \urcorner$). By the choice of A it follows that $T \vdash \neg A$, which contradicts the consistency of T.

Ad $(**)$. By $(*)$ we know $T \nvdash A$. Therefore $A \notin T$, hence $\ulcorner A \urcorner \notin \ulcorner T \urcorner$ and hence $T \vdash \neg B(\underline{\ulcorner A \urcorner})$. By the choice of A it follows that $T \vdash A$. □

THEOREM (Gödel's First Incompleteness Theorem). *Assume that T is an axiomatized consistent theory with the property that all recursive functions are representable in T. Then T is incomplete.*

PROOF. This is an immediate consequence of the above theorem and the corollary in 1.4. □

4.2. Rosser's Form of Gödel's First Incompleteness Theorem. As already mentioned, we now want to sharpen the Incompleteness Theorem, by producing a formula A such that neither A nor $\neg A$ is provable. The original idea is due to Rosser.

THEOREM (Gödel-Rosser). *Let T be an axiomatized consistent $\mathcal{L}$-theory with $0, \mathsf{S}, =$ in $\mathcal{L}$ and $\mathsf{Eq}_{\mathcal{L}} \subseteq T$. Assume that there is a formula $L(x, y)$ – written $x < y$ – such that*

$$T \vdash \forall x. x < \underline{a} \to x = \underline{0} \vee \cdots \vee x = \underline{a - 1}, \tag{19}$$

$$T \vdash \forall x. x = \underline{0} \vee \cdots \vee x = \underline{a} \vee \underline{a} < x. \tag{20}$$

Moreover assume that every elementary function is representable in T. Then we can find a closed formula A such that neither A nor $\neg A$ is provable in T.

PROOF. We first define $\mathsf{Refut}_T \subseteq \mathbb{N} \times \mathbb{N}$ by

$$\mathsf{Refut}_T(d, a) :\leftrightarrow \mathsf{Prf}_T(d, \dot{\neg} a).$$

So Refut_T is elementary, and we have $\mathsf{Refut}_T(d, a)$ if and only if d is a refutation of a in T, i.e., d is a derivation of a sequent $\Gamma \Rightarrow \neg A$ coded by $a = \ulcorner \neg A \urcorner$ and Γ is composed from stability axioms and formulas from Ax_T. Let $B_{\mathsf{Prf}_T}(x_1, x_2)$ and $B_{\mathsf{Refut}_T}(x_1, x_2)$ be formulas representing Prf_T and Refut_T, respectively. Choose by the Fixed Point Lemma in 3.2 a closed formula A such that

$$T \vdash A \leftrightarrow \forall x. B_{\mathsf{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \exists y. y < x \wedge B_{\mathsf{Refut}_T}(y, \underline{\ulcorner A \urcorner}).$$

So A expresses its own underivability, in the form (due to Rosser) "For every proof of me there is a shorter proof of my negation".

We shall show $(*)$ $T \nvdash A$ and $(**)$ $T \nvdash \neg A$. Ad $(*)$. Assume $T \vdash A$. Choose a such that

$$\mathsf{Prf}_T(a, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathsf{Refut}_T(b, \ulcorner A \urcorner) \qquad \text{for all } b,$$

since T is consistent. Hence we have

$$T \vdash B_{\mathsf{Prf}_T}(\underline{a}, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg B_{\mathsf{Refut}_T}(\underline{b}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } b.$$

By (19) we can conclude

$$T \vdash B_{\mathsf{Prf}_T}(\underline{a}, \underline{\ulcorner A \urcorner}) \wedge \forall y. y < \underline{a} \to \neg B_{\mathsf{Refut}_T}(y, \underline{\ulcorner A \urcorner})$$

Hence we have

$$T \vdash \exists x. B_{\mathsf{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \wedge \forall y. y < x \to \neg B_{\mathsf{Refut}_T}(y, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg A.$$

This contradicts the assumed consistency of T.

Ad $(**)$. Assume $T \vdash \neg A$. Choose a such that

$$\mathsf{Refut}_T(a, \ulcorner A \urcorner).$$

Then we also have

$$\text{not } \mathsf{Prf}_T(b, \ulcorner A \urcorner) \qquad \text{for all } b,$$

since T is consistent. Hence we have

$$T \vdash B_{\mathsf{Refut}_T}(\underline{a}, \underline{\ulcorner A \urcorner}),$$
$$T \vdash \neg B_{\mathsf{Prf}_T}(\underline{b}, \underline{\ulcorner A \urcorner}) \qquad \text{for all } b.$$

But this implies

$$T \vdash \forall x. B_{\mathsf{Prf}_T}(x, \underline{\ulcorner A \urcorner}) \to \exists y. y < x \wedge B_{\mathsf{Refut}_T}(y, \underline{\ulcorner A \urcorner}),$$

as can be seen easily by cases on x, using (20). Hence $T \vdash A$. But this again contradicts the assumed consistency of T. □

4.3. Relativized Gödel-Rosser Theorem. Finally we formulate a variant of this theorem which does not assume any more that the theory T talks about numbers only.

THEOREM (Gödel-Rosser). *Assume that T is an axiomatized consistent $\mathcal{L}$-theory with $0, \mathsf{S}, =$ in $\mathcal{L}$ and $\mathsf{Eq}_{\mathcal{L}} \subseteq T$. Furthermore assume that there are formulas $N(x)$ and $L(x,y)$ – written Nx and $x < y$ – such that $T \vdash N0$, $T \vdash \forall x{\in}N\, N(\mathsf{S}(x))$ and*

$$T \vdash \forall x{\in}N.x < \underline{a} \to x = \underline{0} \vee \cdots \vee x = \underline{a-1},$$
$$T \vdash \forall x{\in}N.x = \underline{0} \vee \cdots \vee x = \underline{a} \vee \underline{a} < x.$$

Here $\forall x{\in}N\, A$ is short for $\forall x.Nx \to A$. Moreover assume that every elementary function is representable in T. Then one can find a closed formula A such that neither A nor $\neg A$ is provable in T.

PROOF. As before; just relativize all quantifiers to N. □

5. Representability

We show in this section that already very simple theories have the property that all recursive functions are representable in them.

5.1. A Weak Arithmetic. It is here where the need for Gödel's β-function arises: Recall that we had used it to prove that the class of recursive functions can be generated without use of the primitive recursion scheme, i.e. with composition and the unbounded μ-operator as the only generating schemata.

THEOREM. *Assume that T is an $\mathcal{L}$-theory with $0, \mathsf{S}, =$ in $\mathcal{L}$ and $\mathsf{Eq}_{\mathcal{L}} \subseteq T$. Furthermore assume that there are formulas $N(x)$ and $L(x,y)$ – written Nx and $x < y$ – such that $T \vdash N0$, $T \vdash \forall x{\in}N\, N(\mathsf{S}(x))$ and the following hold:*

(21) $T \vdash \mathsf{S}(\underline{a}) \neq 0$ *for all $a \in \mathbb{N}$,*

(22) $T \vdash \mathsf{S}(\underline{a}) = \mathsf{S}(\underline{b}) \to \underline{a} = \underline{b}$ *for all $a, b \in \mathbb{N}$,*

(23) *the functions $+$ and $\cdot$ are representable in T,*

(24) $T \vdash \forall x{\in}N\, (x \not< 0)$,

(25) $T \vdash \forall x{\in}N.x < \mathsf{S}(\underline{b}) \to x < \underline{b} \vee x = \underline{b}$ *for all $b \in \mathbb{N}$,*

(26) $T \vdash \forall x{\in}N.x < \underline{b} \vee x = \underline{b} \vee \underline{b} < x$ *for all $b \in \mathbb{N}$.*

Here again $\forall x{\in}N\, A$ is short for $\forall x.Nx \to A$. Then T fulfills the assumptions of the theorem in 4.3., i.e., the Gödel-Rosser Theorem relativized to N. In particular we have, for all $a \in \mathbb{N}$

(27) $T \vdash \forall x{\in}N.x < \underline{a} \to x = \underline{0} \vee \cdots \vee x = \underline{a-1}$,

(28) $T \vdash \forall x{\in}N.x = \underline{0} \vee \cdots \vee x = \underline{a} \vee \underline{a} < x$,

and every recursive function is representable in T.

PROOF. (27) can be proved easily by induction on a. The base case follows from (24), and the step from the induction hypothesis and (25). (28) immediately follows from the trichotomy law (26), using (27).

For the representability of recursive functions, first note that the formulas $x = y$ and $x < y$ actually do represent in T the equality and the

less-than relations, respectively. From (21) and (22) we can see immediately that $T \vdash \underline{a} \neq \underline{b}$ when $a \neq b$. Assume $a \not< b$. We show $T \vdash \underline{a} \not< \underline{b}$ by induction on b. $T \vdash \underline{a} \not< 0$ follows from (24). In the step we have $a \not< b+1$, hence $a \not< b$ and $a \neq b$, hence by induction hypothesis and the representability (above) of the equality relation, $T \vdash \underline{a} \not< \underline{b}$ and $T \vdash \underline{a} \neq \underline{b}$, hence by (25) $T \vdash \underline{a} \not< \mathsf{S}(\underline{b})$. Now assume $a < b$. Then $T \vdash \underline{a} \neq \underline{b}$ and $T \vdash \underline{b} \not< \underline{a}$, hence by (26) $T \vdash \underline{a} < \underline{b}$.

We now show by induction on the definition of μ-recursive functions, that every recursive function is representable in T. Observe first that the second condition (17) in the definition of representability of a function automatically follows from the other two conditions (and hence need not be checked further). This is because if $c \neq f(a_1, \ldots, a_n)$ then by contraposing the third condition (18),

$$T \vdash \underline{c} \neq \underline{f(a_1, \ldots, a_n)} \to A(\underline{a_1}, \ldots, \underline{a_n}, \underline{f(a_1, \ldots, a_n)}) \to \neg A(\underline{a_1}, \ldots, \underline{a_n}, \underline{c})$$

and hence by using representability of equality and the first representability condition (16) we obtain $T \vdash \neg A(\underline{a_1}, \ldots, \underline{a_n}, \underline{c})$

The *initial functions* constant 0, successor and projection (onto the i-th coordinate) are trivially represented by the formulas $0 = y$, $\mathsf{S}(x) = y$ and $x_i = y$ respectively. Addition and multiplication are represented in T by assumption. Recall that the one remaining initial function of μ-recursiveness is $\dot{-}$, but this is definable from the characteristic function of $<$ by $a \dot{-} b = \mu i.\, b + i \geq a = \mu i.\, c_<(b+i, a) = 0$. We now show that the characteristic function of $<$ is representable in T. (It will then follow that $\dot{-}$ is representable, once we have shown that the representable functions are closed under μ.) So define

$$A(x_1, x_2, y) := (x_1 < x_2 \wedge y = 1) \vee (x_1 \not< x_2 \wedge y = 0).$$

Assume $a_1 < a_2$. Then $T \vdash \underline{a_1} < \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 1)$. Now assume $a_1 \not< a_2$. Then $T \vdash \underline{a_1} \not< \underline{a_2}$, hence $T \vdash A(\underline{a_1}, \underline{a_2}, 0)$. Furthermore notice that $A(x_1, x_2, y) \wedge A(x_1, x_2, z) \to y = z$ already follows logically from the equality axioms (by cases on $x_1 < x_2$).

For the *composition* case, suppose f is defined from $h, g_1, \ldots, g_m$ by

$$f(\vec{a}) = h(g_1(\vec{a}), \ldots, g_m(\vec{a})).$$

By induction hypothesis we already have representing formulas $A_{g_i}(\vec{x}, y_i)$ and $A_h(\vec{y}, z)$. As representing formula for f we take

$$A_f := \exists \vec{y}. A_{g_1}(\vec{x}, y_1) \wedge \cdots \wedge A_{g_m}(\vec{x}, y_m) \wedge A_h(\vec{y}, z).$$

Assume $f(\vec{a}) = c$. Then there are $b_1, \ldots, b_m$ such that $T \vdash A_{g_i}(\underline{\vec{a}}, \underline{b_i})$ for each i, and $T \vdash A_h(\underline{\vec{b}}, \underline{c})$ so by logic $T \vdash A_f(\underline{\vec{a}}, \underline{c})$. It remains to show uniqueness $T \vdash A_f(\underline{\vec{a}}, z_1) \to A_f(\underline{\vec{a}}, z_2) \to z_1 = z_2$. But this follows by logic from the induction hypothesis for g_i, which gives

$$T \vdash A_{g_i}(\underline{\vec{a}}, y_{1i}) \to A_{g_i}(\underline{\vec{a}}, y_{2i}) \to y_{1i} = y_{2i} = \underline{g_i(\vec{a})}$$

and the induction hypothesis for h, which gives

$$T \vdash A_h(\underline{\vec{b}}, z_1) \to A_h(\underline{\vec{b}}, z_2) \to z_1 = z_2 \quad \text{with } b_i = g_i(\vec{a}).$$

For the μ case, suppose f is defined from g (taken here to be binary for notational convenience) by $f(a) = \mu i\, (g(i, a) = 0)$, assuming $\forall a \exists i\, (g(i, a) =$

0). By induction hypothesis we have a formula $A_g(y, x, z)$ representing g. In this case we represent f by the formula

$$A_f(x, y) := Ny \wedge A_g(y, x, 0) \wedge \forall v{\in}N.v < y \to \exists u.u \neq 0 \wedge A_g(v, x, u).$$

We first show the representability condition (16), that is $T \vdash A_f(\underline{a}, \underline{b})$ when $f(a) = b$. Because of the form of A_f this follows from the assumed representability of g together with $T \vdash v < \underline{b} \to v = \underline{0} \vee \cdots \vee v = \underline{b-1}$.

We now tackle the uniqueness condition (18). Given a, let $b := f(a)$ (thus $g(b, a) = 0$ and b is the least such). It suffices to show $T \vdash A_f(\underline{a}, y) \to y = \underline{b}$, and we do this by proving $T \vdash y < \underline{b} \to \neg A_f(\underline{a}, y)$ and $T \vdash \underline{b} < y \to \neg A_f(\underline{a}, y)$, and then appealing to the trichotomy law.

We first show $T \vdash y < \underline{b} \to \neg A_f(\underline{a}, y)$. Now since, for any $i < b$, $T \vdash \neg A_g(\underline{i}, \underline{a}, 0)$ by the assumed representability of g, we obtain immediately $T \vdash \neg A_f(\underline{a}, \underline{i})$. Hence because of $T \vdash y < \underline{b} \to y = \underline{0} \vee \cdots \vee y = \underline{b-1}$ the claim follows.

Secondly, $T \vdash \underline{b} < y \to \neg A_f(\underline{a}, y)$ follows almost immediately from $T \vdash \underline{b} < y \to A_f(\underline{a}, y) \to \exists u.u \neq 0 \wedge A_g(\underline{b}, \underline{a}, u)$ and the uniqueness for g, $T \vdash A_g(\underline{b}, \underline{a}, u) \to u = 0$. This now completes the proof. □

5.2. Robinson's Theory Q. We conclude this section by considering a special and particularly simple arithmetical theory due originally to Robinson. Let $\mathcal{L}_1$ be the language given by 0, S, $+$, $\cdot$ and $=$, and let Q be the theory determined by the axioms $\mathsf{Eq}_{\mathcal{L}_1}$ and

$$\mathsf{S}(x) \neq 0, \tag{29}$$

$$\mathsf{S}(x) = \mathsf{S}(y) \to x = y, \tag{30}$$

$$x + 0 = x, \tag{31}$$

$$x + \mathsf{S}(y) = \mathsf{S}(x + y), \tag{32}$$

$$x \cdot 0 = 0, \tag{33}$$

$$x \cdot \mathsf{S}(y) = x \cdot y + x, \tag{34}$$

$$\exists z\,(x + \mathsf{S}(z) = y) \vee x = y \vee \exists z\,(y + \mathsf{S}(z) = x). \tag{35}$$

THEOREM. *Every theory $T \supseteq Q$ fulfills the assumptions of the theorem of Gödel-Rosser in 4.3, w.r.t. the definition $L(x, y) := \exists z\,(x + \mathsf{S}(z) = y)$ of the $<$-relation. Moreover, every recursive function is representable in T.*

PROOF. We show that T with $N(x) := (x = x)$ and $L(x, y) := \exists z\,(x + \mathsf{S}(z) = y)$ satisfies the conditions of the theorem in 5.1. For (21) and (22) this is clear. For (23) we can take $x + y = z$ and $x \cdot y = z$ as representing formulas. For (24) we have to show $\neg\exists z\,(x + \mathsf{S}(z) = 0)$; this follows from (32) and (29). For the proof of (25) we need the auxiliary proposition

$$x = 0 \vee \exists y\,(x = 0 + \mathsf{S}(y)), \tag{36}$$

which will be attended to below. So assume $x + \mathsf{S}(z) = \mathsf{S}(\underline{b})$, hence also $\mathsf{S}(x+z) = \mathsf{S}(\underline{b})$ and therefore $x+z = \underline{b}$. We now use (36) for z. In case $z = 0$ we obtain $x = \underline{b}$, and in case $\exists y\,(z = 0 + \mathsf{S}(y))$ we have $\exists y'\,(x + \mathsf{S}(y') = \underline{b})$, since $0 + \mathsf{S}(y) = \mathsf{S}(0 + y)$. Thus (25) is proved. (26) follows immediately from (35).

For the proof of (36) we use (35) with $y = 0$. It clearly suffices to exclude the first case $\exists z\,(x + \mathsf{S}(z) = 0)$. But this means $\mathsf{S}(x + z) = 0$, contradicting (29). □

COROLLARY (Essential Undecidability of Q). *Every consistent theory $T \supseteq Q$ is non-recursive.*

PROOF. By the theorems in 5.2 and 4.1. □

5.3. Undecidability of First Order Logic. As a simple corollary to the (essential) undecidability of Q we even obtain the undecidability of pure logic.

COROLLARY (Undecidability of First Order Logic). *The set of formulas derivable in classical first order logic is non-recursive.*

PROOF. Otherwise Q would be recursive, because a formula A is derivable in Q if and only if the implication $B \to A$ is derivable in classical first order logic, where B is the conjunction of the finitely many axioms and equality axioms of Q. □

REMARK. Notice that it suffices that the first order logic should have one binary relation symbol (for $=$), one constant symbol (for 0), one unary function symbol (for S) and two binary functions symbols (for $+$ and $\cdot$). The study of decidable fragments of first order logic is one of the oldest research areas of Mathematical Logic. For more information see Börger, Grädel and Gurevich [**3**].

5.4. Representability by Σ_1-formulas of the language $\mathcal{L}_1$. By reading through the above proof of representability, one sees easily that the representing formulas used are of a restricted form, having no unbounded universal quantifiers and therefore defining Σ^0_1-relations. This will be of crucial importance for our proof of Gödel's Second Incompleteness Theorem to follow, but in addition we need to make a syntactically precise definition of the class of formulas actually involved.

DEFINITION. The Σ_1-formulas of the language $\mathcal{L}_1$ are those generated inductively by the following clauses:

- Only atomic formulas of the restricted forms $x = y$, $x \neq y$, $0 = x$, $\mathsf{S}(x) = y$, $x + y = z$ and $x \cdot y = z$ are allowed as Σ_1-formulas.
- If A and B are Σ_1-formulas, then so are $A \wedge B$ and $A \vee B$.
- If A is a Σ_1-formula, then so is $\forall x{<}y\,A$, which is an abbreviation for $\forall x.\exists z\,(x + \mathsf{S}(z) = y) \to A$.
- If A is a Σ_1-formula, then so is $\exists x\,A$.

COROLLARY. *Every recursive function is representable in Q by a Σ_1-formula in the language $\mathcal{L}_1$.*

PROOF. This can be seen immediately by inspecting the proof of the theorem in 5.1. Only notice that because of the equality axioms $\exists z\,(x + \mathsf{S}(z) = y)$ is equivalent to $\exists z \exists w\,(\mathsf{S}(z) = w \wedge x + w = y)$ and $A(0)$ is equivalent to $\exists x.0 = x \wedge A$. □

6. Unprovability of Consistency

We have seen in the Gödel-Rosser Theorem how, for every axiomatized consistent theory T safisfying certain weak assumptions, we can construct an undecidable sentence A meaning "For every proof of me there is a shorter proof of my negation". Because A is unprovable, it is clearly true.

Gödel's Second Incompleteness Theorem provides a particularly interesting alternative to A, namely a formula Con_T expressing the consistency of T. Again it turns out to be unprovable and therefore true.

We shall prove this theorem in a sharpened form due to Löb.

6.1. Σ_1-Completeness of Q. We begin with an auxiliary proposition, expressing the completeness of Q with respect to Σ_1-formulas.

LEMMA. *Let $A(x_1, \ldots, x_n)$ be a Σ_1-formulas in the language $\mathcal{L}_1$ determined by 0, S, $+$, $\cdot$ und $=$. Assume that $\mathcal{N}_1 \models A[a_1, \ldots, a_n]$ where $\mathcal{N}_1$ is the standard model of $\mathcal{L}_1$. Then $Q \vdash A(\underline{a_1}, \ldots, \underline{a_n})$.*

PROOF. By induction on the Σ_1-formulas of the language $\mathcal{L}_1$. For atomic formulas, the cases have been dealt with either in the earlier parts of the proof of the theorem in 5.1, or (for $x + y = z$ and $x \cdot y = z$) they follow from the recursion equations (31) - (34).

Cases $A \wedge B$, $A \vee B$. The claim follows immediately from the induction hypothesis.

Case $\forall x{<}y\, A(x, y, z_1, \ldots, z_n)$; for simplicity assume $n = 1$. Suppose $\mathcal{N}_1 \models (\forall x{<}y\, A)[b, c]$. Then also $\mathcal{N}_1 \models A[i, b, c]$ for each $i < b$ and hence by induction hypothesis $Q \vdash A(\underline{i}, \underline{b}, \underline{c})$. Now by the theorem in 5.2

$$Q \vdash \forall x{<}\underline{b}.x = \underline{0} \vee \cdots \vee x = \underline{b-1},$$

hence

$$Q \vdash (\forall x{<}y\, A)(\underline{b}, \underline{c}).$$

Case $\exists x\, A(x, y_1, \ldots, y_n)$; for simplicity take $n = 1$. Assume $\mathcal{N}_1 \models (\exists x\, A)[b]$. Then $\mathcal{N}_1 \models A[a, b]$ for some $a \in \mathbb{N}$, hence by induction hypothesis $Q \vdash A(\underline{a}, \underline{b})$ and therefore $Q \vdash \exists x\, A(x, \underline{b})$. □

6.2. Formalized Σ_1-Completeness.

LEMMA. *In an appropriate theory T of arithmetic with induction, we can formally prove for any Σ_1-formula A*

$$A(\vec{x}) \to \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{\vec{x}}) \urcorner).$$

Here $\mathsf{Prf}_T(p, z)$ is a suitable Σ_1-formula which represents in Robinson's Q the recursive relation "a is the Gödel number of a proof in T of the formula with Gödel number b". Also $\ulcorner A(\dot{x}) \urcorner$ is a term which represents, in Q, the numerical function mapping a number a to the Gödel number of $A(\underline{a})$.

PROOF. We have not been precise about the theory T in which this result is to be formalized, but we shall content ourselves at this stage with merely pointing out, as we proceed, the basic properties that are required. Essentially T will be an extension of Q, together with induction formalized by the axiom schema

$$B(0) \wedge (\forall x.B(x) \to B(\mathsf{S}(x))) \to \forall x\, B(x)$$

and it will be assumed that T has sufficiently many basic functions available to deal with the construction of appropriate Gödel numbers.

The proof goes by induction on the build-up of the Σ_1-formula $A(\vec{x})$.

We consider three atomic cases, leaving the others to the reader. Suppose $A(x)$ is the formula $0 = x$. We show $T \vdash 0 = x \to \exists p \mathsf{Prf}_T(p, \ulcorner 0 = \dot{x} \urcorner)$, by induction on x. The base case merely requires the construction of a numeral representing the Gödel number of the axiom $0 = 0$, and the induction step is trivial because $T \vdash \mathsf{S}(x) \neq 0$. Secondly suppose A is the formula $x + y = z$. We show $T \vdash \forall z. x + y = z \to \exists p \mathsf{Prf}_T(p, \ulcorner \dot{x} + \dot{y} = \dot{z} \urcorner)$ by induction on y. If $y = 0$, the assumption gives $x = z$ and one requires only the Gödel number for the axiom $\forall x(x + 0 = x)$ which, when applied to the Gödel number of the x-th numeral, gives $\exists p \mathsf{Prf}_T(p, \ulcorner \dot{x} + 0 = \dot{z} \urcorner)$. If y is a successor $\mathsf{S}(u)$, then the assumption gives $z = \mathsf{S}(v)$ where $x + u = v$, so by the induction hypothesis we already have a p such that $\mathsf{Prf}_T(p, \ulcorner \dot{x} + \dot{u} = \dot{v} \urcorner)$. Applying the successor to both sides, one then easily obtains from p a p' such that $\mathsf{Prf}_T(p', \ulcorner \dot{x} + \dot{y} = \dot{z} \urcorner)$. Thirdly suppose A is the formula $x \neq y$. We show $T \vdash \forall y. x \neq y \to \exists p \mathsf{Prf}_T(p, \ulcorner \dot{x} \neq \dot{y} \urcorner)$ by induction on x. The base case $x = 0$ requires a subinduction on y. If $y = 0$, then the claim is trivial (by ex-falso). If $y = \mathsf{S}(u)$, we have to produce a Gödel number p such that $\mathsf{Prf}_T(p, \ulcorner 0 \neq \mathsf{S}(\dot{u} \urcorner)$, but this is just an axiom. Now consider the step case $x = \mathsf{S}(v)$. Again we need an auxiliary induction on y. Its base case is dealt with exactly as before, and when $y = \mathsf{S}(u)$ it uses the induction hypothesis for $v \neq u$ together with the injectivity of the successor.

The cases where A is built up by conjunction or disjunction are rather trivial. One only requires, for example in the conjunction case, a function which combines the Gödel numbers of the proofs of the separate conjuncts into a single Gödel number of a proof of the conjunction A itself.

Now consider the case $\exists y A(y, x)$ (with just one parameter x for simplicity). By the induction hypothesis we already have $T \vdash A(y, x) \to \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{y}, \dot{x}) \urcorner)$. But any Gödel number p such that $\mathsf{Prf}_T(p, \ulcorner A(\dot{y}, \dot{x}) \urcorner)$ can easily be transformed (by formally applying the $\exists$-rule) into a Gödel number p' such that $\mathsf{Prf}_T(p', \ulcorner \exists y A(y, \dot{x}) \urcorner)$. Therefore we obtain as required, $T \vdash \exists y A(y, x) \to \exists p' \mathsf{Prf}_T(p', \ulcorner \exists y A(y, \dot{x}) \urcorner)$.

Finally suppose the Σ_1-formula is of the form $\forall u{<}y\, A(u, x)$. We must show

$$\forall u{<}y\, A(u, x) \to \exists p \mathsf{Prf}_T(p, \ulcorner \forall u{<}\dot{y}\, A(u, \dot{x}) \urcorner).$$

By the induction hypothesis

$$T \vdash A(u, x) \to \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{u}, \dot{x}) \urcorner)$$

so by logic

$$T \vdash \forall u{<}y\, A(u, x) \to \forall u{<}y \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{u}, \dot{x}) \urcorner).$$

The required result now follows immediately from the auxiliary lemma:

$$T \vdash \forall u{<}y \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{u}, \dot{x}) \urcorner) \to \exists q \mathsf{Prf}_T(q, \ulcorner \forall u{<}\dot{y}\, A(u, \dot{x}) \urcorner).$$

It remains only to prove this, which we do by induction on y (inside T). In case $y = 0$ a proof of $u < 0 \to A$ is trivial, by ex-falso, so the required Gödel number q is easily constructed. For the step case $y = \mathsf{S}(z)$ by assumption we have $\forall u{<}z \exists p \mathsf{Prf}_T(p, \ulcorner A(\dot{u}, \dot{x}) \urcorner)$, hence $\exists q \mathsf{Prf}_T(q, \ulcorner \forall u{<}\dot{z}\, A(u, \dot{x}) \urcorner)$ by IH.

Also $\exists p' \mathsf{Prf}_T(p', \ulcorner A(\dot{z}, \dot{x}) \urcorner)$. Now we only have to combine p' and q to obtain (by means of an appropriate "simple" function) a Gödel number q' so that $\mathsf{Prf}_T(q', \ulcorner \forall u{<}\dot{y}\, A(u, \dot{x}) \urcorner)$. □

6.3. Derivability Conditions. So now let T be an axiomatized consistent theory with $T \supseteq Q$, and possessing "enough" induction to formalize Σ_1-completeness as we have just done. Define, from the associated formula Prf_T, the following $\mathcal{L}_1$-formulas:

$$\begin{aligned} \mathsf{Thm}_T(x) &:= \exists y\, \mathsf{Prf}_T(y, x), \\ \mathsf{Con}_T &:= \neg \exists y\, \mathsf{Prf}_T(y, \ulcorner \underline{\bot} \urcorner). \end{aligned}$$

So $\mathsf{Thm}_T(x)$ defines in $\mathcal{N}_1$ the set of formulas provable in T, and we have $\mathcal{N}_1 \models \mathsf{Con}_T$ if and only if T is consistent. For $\mathcal{L}_1$-formulas A let $\Box A := \mathsf{Thm}_T(\ulcorner \underline{A} \urcorner)$.

Now consider the following two *derivability conditions* for T (Hilbert-Bernays [**12**])

(37) $\quad T \vdash A \to \Box A \quad (A$ closed Σ_1-formula of the language $\mathcal{L}_1)$,

(38) $\quad T \vdash \Box(A \to B) \to \Box A \to \Box B$.

(37) is just a special case of formalized Σ_1-completeness for closed formulas, and (38) requires only that the theory T has a term that constructs, from the Gödel number of a proof of $A \to B$ and the Gödel number of a proof of A, the Gödel number of a proof of B, and furthermore this fact must be provable in T.

Theorem (Gödel's Second Incompleteness Theorem). *Let T be an axiomatized consistent extension of Q, satisfying the derivability conditions* (37) *und* (38). *Then $T \nvdash \mathsf{Con}_T$.*

Proof. Let $C := \bot$ in the theorem below, which is Löb's generalization of Gödel's original proof. □

Theorem (Löb). *Let T be an axiomatized consistent extension of Q satisfying the derivability conditions* (37) *and* (38). *Then for any closed $\mathcal{L}_1$-formula C, if $T \vdash \Box C \to C$ (that is, $T \vdash \mathsf{Thm}_T(\ulcorner \underline{C} \urcorner) \to C$), then already $T \vdash C$.*

Proof. Assume $T \vdash \Box C \to C$. Choose A by the Fixed Point Lemma in 3.2 such that

(39) $\qquad Q \vdash A \leftrightarrow (\Box A \to C)$.

We must show $T \vdash C$. First we show $T \vdash \Box A \to C$, as follows.

$T \vdash A \to \Box A \to C$	by (39)
$T \vdash \Box(A \to \Box A \to C)$	by Σ_1-completeness
$T \vdash \Box A \to \Box(\Box A \to C)$	by (38)
$T \vdash \Box A \to \Box\Box A \to \Box C$	again by (38)
$T \vdash \Box A \to \Box C$	because $T \vdash \Box A \to \Box\Box A$ by (37).

Therefore from the assumption $T \vdash \Box C \to C$ we obtain $T \vdash \Box A \to C$.

This implies $T \vdash A$ by (39), and then $T \vdash \Box A$ by Σ_1-completeness. But $T \vdash \Box A \to C$ as we have just shown, therefore $T \vdash C$. □

REMARK. It follows immediately that if T is any axiomatized consistent extension of Q satisfying the derivability conditions (37) und (38), then the reflection scheme

$$\mathsf{Thm}_T(\underline{\ulcorner C \urcorner}) \to C \quad \text{for closed } \mathcal{L}_1\text{-formulas } C$$

is not derivable in T. For by Löb's Theorem, it cannot be derivable when C is underivable.

By adding to Q the induction scheme for all formulas we obtain Peano-arithmetic **PA**, which is the most natural example of a theory T to which the results above apply. However, various weaker fragments of **PA**, obtained by restricting the classes of induction formulas, would serve equally well as examples of such T.

7. Notes

The undecidability of first order logic has first been proved by Church; however, the basic idea of the proof was present in Gödels [**11**] already

The fundamental papers on incompleteness are Gödel's [**10**] (from 1930) and [**11**] (from 1931). Gödel also discovered the β-function, which is of central importance for the representation theorem; he made use of the Fixed Point Lemma only implicitely. His first Incompleteness Theorem is based on the formula "I am not provable", a fixed point of $\neg\mathsf{Thm}_T(x)$. For the independence of this proposition from the underlying theory T he had to assume ω-consistency of T. Rosser (1936) proved the sharper result reproduced here, using a formula with the meaning "for every proof of me there is a shorter proof of my negation". The undefinability of the notion of truth has first been proved by Tarski (1939). The arithmetical theories R und Q_0 (in Exercises 46 and 47) are due to R. Robinson (1950). R is essentially undecidable, incomplete and strong enough for Σ_1-completeness; moreover, all recursive relations are representable in R. Q_0 is a very natural theory and in contrast to R finite. Q_0 is minimal in the following sense: if one axiom is deleted, then the resulting theory is not essentially undecidable any more. The first essentially undecidable theory was found by Mostowski and Tarski (1939); when readfing the manuscript, J. Robinson had the idea of treating recursive functions without the scheme of primitive recursion.

Important examples for undecidable theories are (in historic order): Arithmetic of natural numbers (Rosser, 1936), arithmetic of integers (Tarski, Mostowski, 1949), arithmetic of rationals and the theory of ordered fields (J. Robinson 1949), group theory and lattice theory (Tarski 1949). This is in contrast to the following decidable theories: the theory of addition for natural numbers (Pressburger 1929), that of multiplication (Mostowski 1952), the theory of abelian groups (Szmielew 1949), of algebraically closed fields and of boolean algebras (Tarski 1949), the theory of linearly ordered sets (Ehrenfeucht, 1959).

CHAPTER 5

Set Theory

1. Cumulative Type Structures

Set theory can be viewed as a framework within which mathematics can be given a foundation. Here we want to develop set theory as a formal theory within mathematical logic. But first it is necessary to have an intuitive picture of the notion of a set, to be described by the axioms.

1.1. Cantor's Definition. Cantor in 1895 gave the following definition:

> Unter einer "Menge" verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die Elemente von M genannt werden) zu einem Ganzen.

One can try to make this definition more precise, as follows. Let V be the collection of all objects "unserer Anschauung oder unseres Denkens". Let $A(x)$ denote properties of objects x from V. Then one can form the set $\{\, x \mid A(x) \,\}$, the set of all objects x of V with the property $A(x)$. According to Cantor's definition $\{\, x \mid A(x) \,\}$ is again an object in V.

Examples for properties: (1) x is a natural number. (2) x is a set. (3) x is a point, y is a line and x lies on y. (4) y is a set and x is an element of y, shortly: $\mathsf{Set}(y) \wedge x \in y$.

However, Cantor's definition cannot be accepted in its original form, for it leads to contradictions. The most well known is *Russell's antinomy*: Let $x_0 := \{\, x \mid \mathsf{Set}(x) \wedge x \notin x \,\}$. Then

$$x_0 \in x_0 \leftrightarrow \mathsf{Set}(x_0) \wedge x_0 \notin x_0 \leftrightarrow x_0 \notin x_0,$$

for x_0 is a set.

1.2. Shoenfield's Principle. The root for this contradiction is the fact that in Cantor's definition we accept the concept of a finished totality of all sets. However, this is neither necessary nor does it mirror the usual practice of mathematics. It completely suffices to form a set only if all its elements "are available" already. This leads to the concept of a stepwise construction of sets, or more precisely to the *cumulative type structure*: We start with certain "urelements", that form the sets of level 0. Then on an arbitrary level we can form all sets whose elements belong to earlier levels.

If for instance we take as urelements the natural numbers, then $\{27, \{5\}\}$ belongs to level 2.

The following natural questions pose themselves: (1) Which urelements should we choose? (2) How far do the levels reach?

Ad (1). For the purposes of mathematics it is completely sufficient not to assume any urelements at all; then one speaks of *pure sets*. This will be done in the following.

Level 0: $-$
Level 1: $\emptyset$
Level 2: $\emptyset, \{\emptyset\}$
Level 3: $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}$
and so on.

Ad (2). In [**23**], Shoenfield formulated the following principle:

> Consider a collection $\mathcal{S}$ of levels. If a situation can be conceived where all the levels from $\mathcal{S}$ are constructed, then there exists a level which is past all those levels.

From this admittedly rather vage principle we shall draw exact consequences, which will be fixed as axioms.

By a *set* we intuitively understand an object that belongs to some level of the cumulative type structure. By a *class* we mean an arbitrary collection of sets.

So every set clearly is a class. Moreover there are classes that are not sets, for instance the class V of all sets.

2. Axiomatic Set Theory

In set theory – as in any axiomatic theory – we have to explicitely state all used properties, including the "obvious" ones.

2.1. Extensionality, Equality. The language of set theory has a single non-logical symbol, the *element* relation $\in$. So the only atomic formulas are of the form $x \in y$ (x is an element of y). Equality $x = y$ is defined by

$$x = y := \forall z.z \in x \leftrightarrow z \in y.$$

To ensure compatibility of the $\in$-relation with equality we need an axiom:

Axiom (Extensionality).

$$x = y \to x \in z \to y \in z.$$

Remark. If alternatively equality is to be used as a primitive symbol, one must require the equality axioms and in addition

$$(\forall z.z \in x \leftrightarrow z \in y) \to x = y.$$

As classes in our axiomatic theory we only allow *definable* collections of sets. By "definable" we mean definable by a formula in the language of set theory. More precisely: If $A(x)$ is a formula, then

$$\{\, x \mid A(x) \,\}$$

denotes the *class* of all sets x with the property $A(x)$.

Instead of classes we could have used properties or more precisely formulas as well. However, classes allow for a simpler and more suggestive formulation of many of the propositions we want to consider.

If $A(x)$ is the formula $x = x$, then $\{ x \mid A(x) \}$ is called the *all class* or the (set theoretic) *universe*. If $A(x)$ is the formula $x \notin x$, then $\{ x \mid A(x) \}$ is called the *Russell class*.

We now give some definitions that will be used all over in the following. A set b is an element of the class $\{ x \mid A(x) \}$ if $A(b)$ holds:

$$b \in \{ x \mid A(x) \} := A(b).$$

Two classes $\mathcal{A}$, $\mathcal{B}$ are *equal* if they have the same elements:

$$\mathcal{A} = \mathcal{B} := \forall x.x \in \mathcal{A} \leftrightarrow x \in \mathcal{B}.$$

If $\mathcal{A}$ is a class and b a set, then $\mathcal{A}$ and b are called equal if they have the same elements:

$$\mathcal{A} = b := \forall x.x \in \mathcal{A} \leftrightarrow x \in b.$$

In this case we identify the class $\mathcal{A}$ with this set b. Instead of "$\mathcal{A}$ ist set" we also write $\mathcal{A} \in V$. A class $\mathcal{B}$ is an element of a set a (of a class $\mathcal{A}$, resp.) if $\mathcal{B}$ is equal to an element x of a (of $\mathcal{A}$, resp.).

$$\mathcal{B} \in a := \exists x.x \in a \wedge \mathcal{B} = x,$$
$$\mathcal{B} \in \mathcal{A} := \exists x.x \in \mathcal{A} \wedge \mathcal{B} = x.$$

A class $\mathcal{A}$ is a *proper class* if $\mathcal{A}$ is not a set:

$$\mathcal{A} \text{ proper class} := \forall x(x \neq \mathcal{A}).$$

Remark. Every set b is a class, since

$$b = \{ x \mid x \in b \}.$$

The Russell class is a proper class, for if $\{ x \mid x \notin x \} = x_0$, we would have

$$x_0 \in x_0 \leftrightarrow x_0 \notin x_0.$$

So the Russell construction is not an antinomy any more, but simply says that there are sets and (proper) classes.

Let $\mathcal{A}$, $\mathcal{B}$ be classes (proper classes or sets) and $a, b, a_1, \ldots, a_n$ sets. We define

$\{a_1, \ldots, a_n\} := \{ x \mid x = a_1 \vee \cdots \vee x = a_n \}$,	
$\emptyset := \{ x \mid x \neq x \}$	empty class,
$V := \{ x \mid x = x \}$	all class,
$\mathcal{A} \subseteq \mathcal{B} := \forall x.x \in \mathcal{A} \rightarrow x \in \mathcal{B}$	$\mathcal{A}$ is subclass of $\mathcal{B}$,
$\mathcal{A} \subsetneq \mathcal{B} := \mathcal{A} \subseteq \mathcal{B} \wedge \mathcal{A} \neq \mathcal{B}$	$\mathcal{A}$ is proper subclass of $\mathcal{B}$,
$\mathcal{A} \cap \mathcal{B} := \{ x \mid x \in \mathcal{A} \wedge x \in \mathcal{B} \}$	intersection,
$\mathcal{A} \cup \mathcal{B} := \{ x \mid x \in \mathcal{A} \vee x \in \mathcal{B} \}$	union,
$\mathcal{A} \setminus \mathcal{B} := \{ x \mid x \in \mathcal{A} \wedge x \notin \mathcal{B} \}$	difference,
$\bigcup \mathcal{A} := \{ x \mid \exists y.y \in \mathcal{A} \wedge x \in y \}$	big union,
$\bigcap \mathcal{A} := \{ x \mid \forall y.y \in \mathcal{A} \rightarrow x \in y \}$	big intersection,
$\mathcal{P}(\mathcal{A}) := \{ x \mid x \subseteq \mathcal{A} \}$	power class of $\mathcal{A}$.

In particular $a \cup b = \bigcup\{a, b\}$ and $a \cap b = \bigcap\{a, b\}$, and $\bigcap \emptyset$ is the all class. Moreover $\mathcal{P}(\mathcal{A})$ is the class of all subclasses of $\mathcal{A}$ that happen to be sets.

2.2. Pairs, Relations, Functions, Unions. Ordered pairs are defined by means of a little trick due to Kuratowski:

$$(a,b) := \{\, x \mid x = \{a\} \vee x = \{a,b\} \,\} \quad \text{(ordered) pair,}$$

so $(a,b) = \{\{a\},\{a,b\}\}$. To make sure that (a,b) is not the empty class, we have to require axiomatically that $\{a\}$ and $\{a,b\}$ are sets:

Axiom (Pairing).

$$\{x,y\} \textit{ is a set.}$$

In the cumulative type structure the pairing axiom clearly holds, because for any two levels S_1 and S_2 by the Shoenfield principle there must be a level S coming after S_1 and S_2.

Explicitely the pairing axiom is $\forall x \forall y \exists z \forall u. u \in z \leftrightarrow u = x \vee u = y$. In particular it follows that for every set a the singleton class $\{a\}$ is a set. It also follows that $(a,b) = \{\{a\},\{a,b\}\}$ is a set.

Moreover we define

$$\{\, (x,y) \mid A(x,y) \,\} := \{\, z \mid \exists x,y. A(x,y) \wedge z = (x,y) \,\}$$

and

$\mathcal{A} \times \mathcal{B} := \{\, (x,y) \mid x \in \mathcal{A} \wedge y \in \mathcal{B} \,\}$	cartesian product of $\mathcal{A}$, $\mathcal{B}$,
$\mathsf{dom}(\mathcal{A}) := \{\, x \mid \exists y\, ((x,y) \in \mathcal{A}) \,\}$	domain of $\mathcal{A}$,
$\mathsf{rng}(\mathcal{A}) := \{\, y \mid \exists x\, ((x,y) \in \mathcal{A}) \,\}$	range of $\mathcal{A}$,
$\mathcal{A}{\restriction}\mathcal{B} := \{\, (x,y) \mid (x,y) \in \mathcal{A} \wedge x \in \mathcal{B} \,\}$	restriction of $\mathcal{A}$ to $\mathcal{B}$,
$\mathcal{A}[\mathcal{B}] := \{\, y \mid \exists x. x \in \mathcal{B} \wedge (x,y) \in \mathcal{A} \,\}$	image of $\mathcal{B}$ under $\mathcal{A}$,
$\mathcal{A}^{-1} := \{\, (y,x) \mid (x,y) \in \mathcal{A} \,\},$	inverse of $\mathcal{A}$,
$\mathcal{A} \circ \mathcal{B} := \{\, (x,z) \mid \exists y. (x,y) \in \mathcal{B} \wedge (y,z) \in \mathcal{A} \,\}$	composition of $\mathcal{A}$, $\mathcal{B}$.

Without any difficulty we can introduce the usual notions concerning relations and functions. For classes $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ we define

(a) $\mathcal{A}$ is a *relation* iff $\mathcal{A} \subseteq V \times V$. Hence a relation is a class of pairs. Instead of $(a,b) \in \mathcal{A}$ we also write $a\mathcal{A}b$.
(b) $\mathcal{A}$ is a *relation on* $\mathcal{B}$ iff $\mathcal{A} \subseteq \mathcal{B} \times \mathcal{B}$.
(c) $\mathcal{A}$ is a *function* iff $\mathcal{A}$ is a relation and

$$\forall x,y,z. (x,y) \in \mathcal{A} \to (x,z) \in \mathcal{A} \to y = z.$$

(d) $\mathcal{A}\colon \mathcal{B} \to \mathcal{C}$ iff $\mathcal{A}$ is a function such that $\mathsf{dom}(\mathcal{A}) = \mathcal{B}$ and $\mathcal{A}[\mathcal{B}] \subseteq \mathcal{C}$. We then call $\mathcal{A}$ a *function from* $\mathcal{B}$ *to* $\mathcal{C}$.
(e) $\mathcal{A}\colon \mathcal{B} \to_{\text{onto}} \mathcal{C}$ iff $\mathcal{A}\colon \mathcal{B} \to \mathcal{C}$ and $\mathcal{A}[\mathcal{B}] = \mathcal{C}$. We then call $\mathcal{A}$ a *surjective function from* $\mathcal{B}$ *onto* $\mathcal{C}$.
(f) $\mathcal{A}$ is *injective* iff $\mathcal{A}$ and $\mathcal{A}^{-1}$ are functions.
(g) $\mathcal{A}\colon \mathcal{B} \leftrightarrow \mathcal{C}$ iff $\mathcal{A}\colon \mathcal{B} \to_{\text{onto}} \mathcal{C}$ and $\mathcal{A}$ is injective. Then $\mathcal{A}$ is called *bijective function from* $\mathcal{B}$ *onto* $\mathcal{C}$.

For the further development of set theory more axioms are necessary, in particular

Axiom (Union).

$$\bigcup x \textit{ is a set.}$$

The union axiom holds in the cumulative type structure. To see this, consider a level S where x is formed. An arbitrary element $v \in x$ then is available at an earlier level S_v already. Similarly every element $u \in v$ is present at a level $S_{v,u}$ before S_v. But all these u make up $\bigcup x$. Hence also $\bigcup x$ can be formed at level S.

Explicitely the union axiom is $\forall x \exists y \forall z. z \in y \leftrightarrow \exists u. u \in x \wedge z \in u$.

We now can extend the previous definition by

$$\mathcal{A}(x) := \bigcup \{\, y \mid (x, y) \in \mathcal{A} \,\} \quad \text{application.}$$

If $\mathcal{A}$ is a function and $(x, y) \in \mathcal{A}$, then $\mathcal{A}(x) = \bigcup\{y\} = y$ and we write $\mathcal{A} \colon x \mapsto y$.

2.3. Separation, Power Set, Replacement Axioms.

Axiom (Separation). *For every class $\mathcal{A}$,*

$$\mathcal{A} \subseteq x \to \exists y\, (\mathcal{A} = y).$$

So the separation scheme says that every subclass $\mathcal{A}$ of a set x is a set. It is valid in the cumulative type structure, since on the same level where x is formed we can also form the set y, whose elements are just the elements of the class $\mathcal{A}$.

Notice that the separation scheme consists of infinitely many axioms.

Axiom (Power set).

$$\mathcal{P}(x) \textit{ is a set.}$$

The power set axiom holds in the cumulative type structure. To see this, consider a level S where x is formed. Then also every subset $y \subseteq x$ has been formed at level S. On the next level S' (which exists by the Shoenfield principle) we can form $\mathcal{P}(x)$.

Explicitely the power set axiom is $\forall x \exists y \forall z. z \in y \leftrightarrow z \subseteq x$.

Lemma 2.1. *$a \times b$ is a set.*

Proof. We show $a \times b \subseteq \mathcal{P}(\mathcal{P}(a \cup b))$. So let $x \in a$ and $y \in b$. Then

$$\begin{aligned} \{x\}, \{x, y\} &\subseteq a \cup b \\ \{x\}, \{x, y\} &\in \mathcal{P}(a \cup b) \\ \{\, \{x\}, \{x, y\} \,\} &\subseteq \mathcal{P}(a \cup b) \\ (x, y) = \{\, \{x\}, \{x, y\} \,\} &\in \mathcal{P}(\mathcal{P}(a \cup b)) \end{aligned}$$

The claim now follows from the union axiom, the pairing axiom, the power set axiom and the separation scheme. □

Axiom (Replacement). *For every class $\mathcal{A}$,*

$$\mathcal{A} \textit{ is a function } \to \forall x\, (\mathcal{A}[x] \textit{ is a set}).$$

Also the replacement scheme holds in the cumulative type structure; however, this requires some more thought. Consider all elements u of the set $x \cap \mathsf{dom}(\mathcal{A})$. For every such u we know that $\mathcal{A}(u)$ is a set, hence is formed at a level S_u of the cumulative type structure. Because $x \cap \mathsf{dom}(\mathcal{A})$ is a set, we can imagine a situation where all S_u for $u \in x \cap \mathsf{dom}(\mathcal{A})$ are constructed. Hence by the Shoenfield principle there must be a level S coming after all these S_u. In S we can form $\mathcal{A}[x]$.

LEMMA 2.2. *The replacement scheme implies the separation scheme.*

PROOF. Let $\mathcal{A} \subseteq x$ and $\mathcal{B} := \{ (u, v) \mid u = v \wedge u \in \mathcal{A} \}$. Then $\mathcal{B}$ is a function and we have $\mathcal{B}[x] = \mathcal{A}$. □

This does not yet conclude our list of axioms of set theory: later we will require the infinity axiom, the regularity axiom and the axiom of choice.

3. Recursion, Induction, Ordinals

We want to develop a general framework for recursive definitions and inductive proofs. Both will be done by means of so-called well-founded relations. To carry this through, we introduce as an auxiliary notion that of a transitively well-founded relation; later we will see that it is equivalent to the notion of a well-founded relation. We then define the natural numbers in the framework of set theory, and will obtain induction and recursion on the natural numbers as special cases of the corresponding general theorems for transitively well-founded relations. By recursion on natural numbers we can then define the transitive closure of a set, and by means of this notion we will be able to show that well-founded relations coincide with the transitively well-founded relations.

Then we study particular well-founded relations. We first show that arbitrary classes together with the $\in$-relation are up to isomorphism the only well-founded extensional relations (Isomorphy Theorem of Mostowski). Then we consider linear well-founded orderings, called well-orderings. Since they will always be extensional, they must be isomorphic to certain classes with the $\in$-relation, which will be called ordinal classes. Ordinals can then be defined as those ordinal classes that happen to be sets.

3.1. Recursion on Transitively Well-Founded Relations. Let $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ denote classes. For an arbitrary relation $\mathcal{R}$ on $\mathcal{A}$ we define

(a) $\hat{x}^{\mathcal{R}} := \{ y \mid y\mathcal{R}x \}$ is the class of $\mathcal{R}$-*predecessors* of x. We shall write $\hat{x}$ instead of $\hat{x}^{\mathcal{R}}$, if $\mathcal{R}$ is clear from the context.

(b) $\mathcal{B} \subseteq \mathcal{A}$ is called $\mathcal{R}$-*transitive* if

$$\forall x.x \in \mathcal{B} \to \hat{x} \subseteq \mathcal{B}.$$

Hence $\mathcal{B} \subseteq \mathcal{A}$ is $\mathcal{R}$-transitive iff $y\mathcal{R}x$ and $x \in \mathcal{B}$ imply $y \in \mathcal{B}$.

(c) Let $\mathcal{B} \subseteq \mathcal{A}$. $x \in \mathcal{B}$ is an $\mathcal{R}$-*minimal element* of $\mathcal{B}$ if $\hat{x} \cap \mathcal{B} = \emptyset$.

(d) $\mathcal{R}$ is a *transitively well-founded relation* on $\mathcal{A}$ if

(i) Every nonempty subset of $\mathcal{A}$ has an $\mathcal{R}$-minimal element, i.e.

$$\forall a.a \subseteq \mathcal{A} \to a \neq \emptyset \to \exists x.x \in a \wedge \hat{x} \cap a = \emptyset.$$

(ii) For every $x \in \mathcal{A}$ there is an $\mathcal{R}$-transitive set $b \subseteq \mathcal{A}$ such that $\hat{x} \subseteq b$.

We shall almost everywhere omit $\mathcal{R}$, if $\mathcal{R}$ is clear from the context.

REMARK. Let $\mathcal{R}$ be a relation on $\mathcal{A}$. $\mathcal{R}$ is a *transitive relation on* $\mathcal{A}$ if for all $x, y, z \in \mathcal{A}$

$$x\mathcal{R}y \to y\mathcal{R}z \to x\mathcal{R}z.$$

We have the following connection to the notion of $\mathcal{R}$-transitivity for classes: Let $\mathcal{R}$ be a relation on $\mathcal{A}$. Then

$\mathcal{R}$ is a transitive relation on $\mathcal{A} \leftrightarrow$ for every $y \in \mathcal{A}$, $\hat{y}$ is $\mathcal{R}$-transitive.

PROOF. $\rightarrow$. Let $\mathcal{R}$ be a transitive relation on $\mathcal{A}$, $y \in \mathcal{A}$ and $x \in \hat{y}$, hence $x\mathcal{R}y$. We must show $\hat{x} \subseteq \hat{y}$. So let $z\mathcal{R}x$. We must show $z\mathcal{R}y$. But this follows from the transitivity of $\mathcal{R}$. $\leftarrow$. Let $x, y, z \in \mathcal{A}$, $x\mathcal{R}y$ and $y\mathcal{R}z$. We must show $x\mathcal{R}z$. We have $x\mathcal{R}y$ and $y \in \hat{z}$. Since $\hat{z}$ is $\mathcal{R}$-transitive, we obtain $x \in \hat{z}$, hence $x\mathcal{R}z$. □

LEMMA 3.1. *Let $\mathcal{R}$ be a transitively well-founded relation on $\mathcal{A}$. Then*

(a) *Every nonempty subclass $\mathcal{B} \subseteq \mathcal{A}$ has an $\mathcal{R}$-minimal element.*
(b) *$\forall x.x{\in}\mathcal{A} \rightarrow \hat{x}$ is a set.*

PROOF. (a). Let $\mathcal{B} \subseteq \mathcal{A}$ and $z \in \mathcal{B}$. We may assume that z is not $\mathcal{B}$-minimal, i.e. $\hat{z} \cap \mathcal{B} \neq \emptyset$. By part (ii) of the definition of transitively well-founded relations there exists an $\mathcal{R}$-transitive superset $b \subseteq \mathcal{A}$ of $\hat{z}$. Because of $\hat{z} \cap \mathcal{B} \neq \emptyset$ we have $b \cap \mathcal{B} \neq \emptyset$. By part (i) of the same definition there exists an $\mathcal{R}$-minimal $x \in b \cap \mathcal{B}$, i.e., $\hat{x} \cap b \cap \mathcal{B} = \emptyset$. Since b is $\mathcal{R}$-transitive, from $x \in b$ we obtain $\hat{x} \subseteq b$. Therefore $\hat{x} \cap \mathcal{B} = \emptyset$ and hence x is an $\mathcal{R}$-minimal element of $\mathcal{B}$.

(b). This is a consequence of the separation scheme. □

3.2. Induction and Recursion Theorems. We write $\forall x{\in}\mathcal{A}\ldots$ for $\forall x.x \in \mathcal{A} \rightarrow \ldots$ and similarly $\exists x{\in}\mathcal{A}\ldots$ for $\exists x.x \in \mathcal{A} \wedge \ldots$.

THEOREM 3.2 (Induction Theorem). *Let $\mathcal{R}$ be a transitively well-founded relation on $\mathcal{A}$ and $\mathcal{B}$ an arbitrary class. Then*

$$\forall x{\in}\mathcal{A}.\hat{x} \subseteq \mathcal{B} \rightarrow x \in \mathcal{B}$$

implies $\mathcal{A} \subseteq \mathcal{B}$.

PROOF. Assume $\mathcal{A} \setminus \mathcal{B} \neq \emptyset$. Let x be a minimal element of $\mathcal{A} \setminus \mathcal{B}$. It suffices to show $\hat{x} \subseteq \mathcal{B}$, for then by assumption we obtain $x \in \mathcal{B}$, hence a contradiction. Let $z \in \hat{x}$. By the choice of x we have $z \notin \mathcal{A} \setminus \mathcal{B}$, hence $z \in \mathcal{B}$ (because $z \in \mathcal{A}$ holds, since $\mathcal{R}$ is a relation on $\mathcal{A}$). □

THEOREM 3.3 (Recursion Theorem). *Let $\mathcal{R}$ be a transitively well-founded relation on $\mathcal{A}$ and $\mathcal{G}\colon V \rightarrow V$. Then there exists exactly one function $\mathcal{F}\colon \mathcal{A} \rightarrow V$ such that*

$$\forall x{\in}\mathcal{A}\big(\mathcal{F}(x) = \mathcal{G}(\mathcal{F}{\restriction}\hat{x})\big).$$

PROOF. First observe that for $\mathcal{F}\colon \mathcal{A} \rightarrow V$ we have $\mathcal{F}{\restriction}\hat{x} \subseteq \hat{x} \times \mathcal{F}[\hat{x}]$, hence $\mathcal{F}{\restriction}\hat{x}$ is a set.

Uniqueness. Given $\mathcal{F}_1$, $\mathcal{F}_2$. Consider

$$\{\, x \mid x \in \mathcal{A} \wedge \mathcal{F}_1(x) = \mathcal{F}_2(x) \,\} =: \mathcal{B}.$$

By the Induction Theorem it suffices to show $\forall x{\in}\mathcal{A}.\hat{x} \subseteq \mathcal{B} \rightarrow x \in \mathcal{B}$. So let $x \in \mathcal{A}$ and $\hat{x} \subseteq \mathcal{B}$. Then

$$\begin{aligned}
&\mathcal{F}_1{\restriction}\hat{x} = \mathcal{F}_2{\restriction}\hat{x}\\
&\mathcal{G}\,(\mathcal{F}_1{\restriction}\hat{x}) = \mathcal{G}\,(\mathcal{F}_2{\restriction}\hat{x})\\
&\mathcal{F}_1(x) = \mathcal{F}_2(x)\\
&x \in \mathcal{B}.
\end{aligned}$$

Existence. Let

$$\mathcal{B} := \{\, f \mid f \text{ function, } \mathsf{dom}(f) \text{ } \mathcal{R}\text{-transitive subset of } \mathcal{A}, \\ \forall x{\in}\mathsf{dom}(f)\,(f(x) = \mathcal{G}(f{\restriction}\hat{x}))\,\}$$

and

$$\mathcal{F} := \bigcup \mathcal{B}.$$

We first show that

$$f, g \in \mathcal{B} \to x \in \mathsf{dom}(f) \cap \mathsf{dom}(g) \to f(x) = g(x).$$

So let $f, g \in \mathcal{B}$. We prove the claim by induction on x, i.e., by an application of the Induction Theorem to

$$\{\, x \mid x \in \mathsf{dom}(f) \cap \mathsf{dom}(g) \to f(x) = g(x)\,\}.$$

So let $x \in \mathsf{dom}(f) \cap \mathsf{dom}(g)$. Then

$$\begin{array}{ll} \hat{x} \subseteq \mathsf{dom}(f) \cap \mathsf{dom}(g), & \text{for } \mathsf{dom}(f), \mathsf{dom}(g) \text{ are } \mathcal{R}\text{-transitive} \\ f{\restriction}\hat{x} = g{\restriction}\hat{x} & \text{by IH} \\ \mathcal{G}\,(f{\restriction}\hat{x}) = \mathcal{G}\,(g{\restriction}\hat{x}) & \\ f(x) = g(x). & \end{array}$$

Therefore $\mathcal{F}$ is a function. Now this immediately implies $f \in \mathcal{B} \to x \in \mathsf{dom}(f) \to \mathcal{F}(x) = f(x)$; hence we have shown

$$\mathcal{F}(x) = \mathcal{G}\,(\mathcal{F}{\restriction}\hat{x}) \quad \text{for all } x \in \mathsf{dom}(\mathcal{F}). \tag{40}$$

We now show

$$\mathsf{dom}(\mathcal{F}) = \mathcal{A}.$$

$\subseteq$ is clear. $\supseteq$. Use the Induction Theorem. Let $\hat{y} \subseteq \mathsf{dom}(\mathcal{F})$. We must show $y \in \mathsf{dom}(\mathcal{F})$. This is proved indirectly; so assume $y \notin \mathsf{dom}(\mathcal{F})$. Let b be $\mathcal{R}$-transitive such that $\hat{y} \subseteq b \subseteq \mathcal{A}$. Define

$$g := \mathcal{F}{\restriction}b \cup \{\, (y, \mathcal{G}(\mathcal{F}{\restriction}\hat{y}))\,\}.$$

It clearly suffices to show $g \in \mathcal{B}$, for because of $y \in \mathsf{dom}(g)$ this implies $y \in \mathsf{dom}(\mathcal{F})$ and hence the desired contradiction.

g is a function: This is clear, since $y \notin \mathsf{dom}(\mathcal{F})$ by assumption.

$\mathsf{dom}(g)$ is $\mathcal{R}$-transitive: We have $\mathsf{dom}(g) = (b \cap \mathsf{dom}(\mathcal{F})) \cup \{y\}$. First notice that $\mathsf{dom}(\mathcal{F})$ as a union of $\mathcal{R}$-transitive sets is $\mathcal{R}$-transitive itself. Moreover, since b is $\mathcal{R}$-transitive, also $b \cap \mathsf{dom}(\mathcal{F})$ is $\mathcal{R}$-transitive. Now let $z\mathcal{R}x$ and $x \in \mathsf{dom}(g)$. We must show $z \in \mathsf{dom}(g)$. In case $x \in b \cap \mathsf{dom}(\mathcal{F})$ also $z \in b \cap \mathsf{dom}(\mathcal{F})$ (since $b \cap \mathsf{dom}(\mathcal{F})$ is $\mathcal{R}$-transitive, as we just observed), hence $z \in \mathsf{dom}(g)$. In case $x = y$ we have $z \in \hat{y}$, hence $z \in b$ and $z \in \mathsf{dom}(\mathcal{F})$ be the choice of b and y, hence again $z \in \mathsf{dom}(g)$.

$\forall x{\in}\mathsf{dom}(g)\,\big(g(x) = \mathcal{G}(g{\restriction}\hat{x})\big)$: In case $x \in b \cap \mathsf{dom}(\mathcal{F})$ we have

$$\begin{array}{ll} g(x) = \mathcal{F}(x) & \\ \quad = \mathcal{G}(\mathcal{F}{\restriction}\hat{x}) & \text{by (40)} \\ \quad = \mathcal{G}(g{\restriction}\hat{x}) & \text{since } \hat{x} \subseteq b \cap \mathsf{dom}(\mathcal{F}), \text{ for } b \cap \mathsf{dom}(\mathcal{F}) \text{ is } \mathcal{R}\text{-transitive.} \end{array}$$

In case $x = y$ is $g(x) = \mathcal{G}(\mathcal{F}{\restriction}\hat{x}) = \mathcal{G}(g{\restriction}\hat{x})$, for $\hat{x} = \hat{y} \subseteq b \cap \mathsf{dom}(\mathcal{F})$ by the choice of y. $\square$

3.3. Natural Numbers. Zermelo defined the natural numbers within set theory as follows: $0 = \emptyset$, $1 = \{\emptyset\}$, $2 = \{\{\emptyset\}\}$, $3 = \{\{\{\emptyset\}\}\}$ and so on. A disadvantage of this definition is that it cannot be generalized to the transfinite. Later, John von Neumann proposed to represent the number n by a certain set consisting of exactly n elements, namely

$$n := \{\, 0, 1, \ldots, n-1 \,\}.$$

So $0 = \emptyset$ and $n+1 = \{\, 0, 1, \ldots, n \,\} = \{\, 0, 1, \ldots, n-1 \,\} \cup \{n\}$. Generally we define

$$0 := \emptyset, \quad x + 1 := x \cup \{x\}.$$

In particular, $1 := 0 + 1$, $2 := 1 + 1$, $3 := 2 + 1$ and so on.

In order to know that the class of all natural numbers constructed in this way is a set, we need another axiom:

Axiom (Infinity).

$$\exists x.\emptyset \in x \wedge \forall y.y \in x \to y \cup \{y\} \in x.$$

The Infinity Axiom holds in the cumulative type structure. To see this, observe that $0 = \emptyset$, $1 := \emptyset \cup \{\emptyset\}$, $2 := 1 \cup \{1\}$ and so on are formed at levels S_0, S_1, S_2 ..., and we can conceive a situation where all these levels are completed. By the Shoenfield principle there must be a level - call it S_ω - which is past all these levels. At S_ω we can form ω.

We call a class $\mathcal{A}$ *inductive* if

$$\emptyset \in \mathcal{A} \wedge \forall y.y \in \mathcal{A} \to y \cup \{y\} \in \mathcal{A}.$$

So the infinity axiom says that there is an inductive set. Define

$$\omega := \bigcap \{\, x \mid x \text{ is inductive} \,\}.$$

Clearly ω is a set, with the properties $0 \in \omega$ and $y \in \omega \to y + 1 \in \omega$. ω is called the set of *natural numbers*.

Let n, m denote natural numbers. $\forall n\, A(n)$ is short for $\forall x.x \in \omega \to A(x)$, similarly $\exists n\, A(n)$ for $\exists x.x \in \omega \wedge A(x)$ and $\{\, n \mid A(n) \,\}$ for $\{\, x \mid x \in \omega \wedge A(x) \,\}$.

Theorem 3.4 (Induction on ω).

(a) $x \subseteq \omega \to 0 \in x \to (\forall n.n \in x \to n + 1 \in x) \to x = \omega$.
(b) *For every formula* $A(x)$,

$$A(0) \to (\forall n.A(n) \to A(n+1)) \to \forall n A(n).$$

Proof. (a). x is inductive, hence $\omega \subseteq x$. (b). Let $\mathcal{A} := \{\, n \mid A(n) \,\}$. Then $\mathcal{A} \subseteq \omega$ (so $\mathcal{A}$ is set), and by assumption

$$0 \in \mathcal{A},$$
$$n \in \mathcal{A} \to n + 1 \in \mathcal{A}.$$

By (a), $\mathcal{A} = \omega$. □

We now show that for natural numbers the relation $\in$ has all the properties of $<$, and the relation $\subseteq$ all the properties of $\leq$.

A class $\mathcal{A}$ is called *transitive* if it is $\mathcal{E}$-transitive w.r.t. the special relation $\mathcal{E} := \{\, (x, y) \mid x \in y \,\}$ on V, i.e., if $\forall x.x \in \mathcal{A} \to x \subseteq \mathcal{A}$. Therefore $\mathcal{A}$ is transitive iff

$$y \in x \in \mathcal{A} \to y \in \mathcal{A}.$$

LEMMA 3.5. (a) *n is transitive.*
(b) *ω is transitive.*

PROOF. (a). Induction by n. 0 is transitive. $n \to n+1$. By IH, n is transitive. We must show that $n+1$ is transitive. We argue as follows:

$$\begin{aligned} &y \in x \in n+1 \\ &y \in x \in n \cup \{n\} \\ &y \in x \in n \vee y \in x = n \\ &y \in n \vee y \in n \\ &y \in n \cup \{n\} = n+1. \end{aligned}$$

(b). We show $\forall x. x \in n \to x \in \omega$, by induction on n. 0: Clear. $n \to n+1$. By IH we have $\forall x. x \in n \to x \in \omega$. So assume $x \in n+1$. Then $x \in n \vee x = n$, hence $x \in \omega$. □

LEMMA 3.6. $n \notin n$.

PROOF. Induction on n. 0. Clear. $n \to n+1$: By IH is $n \notin n$. Assume

$$\begin{aligned} &n+1 \in n+1 \\ &n+1 \in n \vee n+1 = n \\ &n \in n+1 \in n \vee n \in n+1 = n \\ &n \in n && \text{for } n \text{ is transitive by Lemma 3.5.} \end{aligned}$$

This is a contradiction to the IH. □

LEMMA 3.7. (a) $n \subseteq m+1 \leftrightarrow n \subseteq m \vee n = m+1$.
(b) $n \subseteq m \leftrightarrow n \in m \vee n = m$.
(c) $n \subseteq m \vee m \subseteq n$.
(d) $n \in m \vee n = m \vee m \in n$.

PROOF. (a). $\leftarrow$ follows from $m \subseteq m+1$. $\rightarrow$. Assume $n \subseteq m+1$. **Case** $m \in n$. We show $n = m+1$. $\subseteq$ holds by assumption. $\supseteq$.

$$\begin{aligned} &p \in m+1 \\ &p \in m \vee p = m \\ &p \in n. \end{aligned}$$

Case $m \notin n$. We show $n \subseteq m$.

$$\begin{aligned} &p \in n \\ &p \in m+1 \\ &p \in m \vee p = m, \end{aligned}$$

but $p = m$ is impossible because of $m \notin n$.

(b). $\leftarrow$ follows from transitivity of m. $\rightarrow$. Induction on m. 0. Clear. $m \to m+1$.

$$\begin{aligned} &n \subseteq m+1 \\ &n \subseteq m \vee n = m+1 && \text{by (a)} \\ &n \in m \vee n = m \vee n = m+1 && \text{by IH} \\ &n \in m+1 \vee n = m+1. \end{aligned}$$

(c). Induction on n. 0. Clear. $n \to n+1$: **Case** $m \subseteq n$. Clear. **Case** $n \subseteq m$. Then

$$\begin{aligned} & n \in m \vee n = m && \text{by (b)} \\ & n, \{n\} \subseteq m \vee m \subseteq n+1 \\ & n+1 \subseteq m \vee m \subseteq n+1 \end{aligned}$$

(d). Follows from (c) and (b). □

THEOREM 3.8 (Peano-Axioms). (a) $n+1 \neq \emptyset$.
(b) $n+1 = m+1 \to n = m$.
(c) $x \subseteq \omega \to 0 \in x \to (\forall n.n \in x \to n+1 \in x) \to x = \omega$.

PROOF. (a). Clear. (c). This is Theorem 3.4(a). (b).

$$\begin{aligned} & n+1 = m+1 \\ & n \in m+1 \wedge m \in n+1 \\ & (n \in m \wedge m \in n) \vee n = m \\ & n \in n \vee n = m \\ & n = m. \end{aligned}$$

This concludes the proof. □

We now treat different forms of induction.

THEOREM 3.9 (Course-of-values induction on ω).
(a) $x \subseteq \omega \to [\forall n.(\forall m.m \in n \to m \in x) \to n \in x] \to x = \omega$.
(b) $[\forall n.(\forall m.m \in n \to A(m)) \to A(n)] \to \forall n A(n)$.

PROOF. (b). Assume $\forall n.(\forall m.m \in n \to A(m)) \to A(n)$; we shall say in this case that $A(n)$ is *progressive*. We show $\forall m.m \in n \to A(m)$, by induction on n. 0. Clear. $n \to n+1$. By IH $\forall m.m \in n \to A(m)$. So let $m \in n+1$. Then $m \in n \vee m = n$. In case $m \in n$ we obtain $A(m)$ by IH, and in case $m = n$ we can infer $A(n)$ from the progressiveness of A, using the IH.

(a). From (b), with $A(y) := y \in x$. □

THEOREM 3.10 (Principle of least element for ω).
(a) $\emptyset \neq x \subseteq \omega \to \exists n.n \in x \wedge n \cap x = \emptyset$.
(b) $\exists n A(n) \to \exists n.A(n) \wedge \neg \exists m.m \in n \wedge A(m)$.

PROOF. (b). By Theorem 3.9(b)

$$[\forall n.(\forall m.m \in n \to \neg A(m)) \to \neg A(n)] \to \forall n \neg A(n).$$

Contraposition gives

$$\begin{aligned} & \exists n A(n) \to \exists n.A(n) \wedge \forall m.m \in n \to \neg A(m) \\ & \exists n A(n) \to \exists n.A(n) \wedge \neg \exists m.m \in n \wedge A(m). \end{aligned}$$

(a). From (b), using $A(y) := y \in x$. □

We now consider recursion on natural numbers, which can be treated as a special case of the Recursion Theorem 3.3. To this end, we identify $\in$ with the relation $\mathcal{E} = \{ (x,y) \mid x \in y \}$ and prove the following lemma:

LEMMA 3.11. $\in \cap(\omega \times \omega)$ *is a transitively well-founded relation on* ω.

PROOF. We show both conditions, from the definition of transitively well-founded relations. (i). Let $\emptyset \neq a \subseteq \omega$. We must show $\exists n.n \in a \wedge n \cap a = \emptyset$. But this is the above principle of the least element. (ii). Clear, since n is transitive. □

THEOREM 3.12 (Course-of-values recursion on ω). *Let $\mathcal{G}\colon V \to V$. Then there is exactly one function $f\colon \omega \to V$ such that*

$$\forall n\,(f(n) = \mathcal{G}(f\restriction n)).$$

PROOF. By the Recursion Theorem 3.3 there is a unique $\mathcal{F}\colon \omega \to V$ such that $\forall n\,(\mathcal{F}(n) = \mathcal{G}(\mathcal{F}\restriction n))$. By Replacement, $\mathsf{rng}(\mathcal{F}) = \mathcal{F}[\omega]$ is a set. By Lemma 2.1 and Separation, also $\mathcal{F} \subseteq \omega \times \mathcal{F}[\omega]$ is a set. □

COROLLARY 3.13. *Let $\mathcal{G}\colon V \to V$ and a be a set. Then there is exactly one function $f\colon \omega \to V$ such that*

$$\begin{aligned} &f(0) = a,\\ &\forall n\,(f(n+1) = \mathcal{G}(f(n))). \end{aligned}$$

PROOF. First observe that $\bigcup(n+1) = n$, because of

$$\begin{aligned} x \in \bigcup(n+1) &\leftrightarrow \exists y.x \in y \in n+1\\ &\leftrightarrow \exists m.x \in m \in n+1\\ &\leftrightarrow \exists m.x \in m \subseteq n\\ &\leftrightarrow x \in n. \end{aligned}$$

For the given $\mathcal{G}$ we will construct $\mathcal{G}'$ such that $\mathcal{G}'(f\restriction n+1) = \mathcal{G}(f(n))$. We define a function $\mathcal{G}'\colon V \to V$ satisfying

$$\mathcal{G}'(x) = \begin{cases} \mathcal{G}(x(\bigcup \mathsf{dom}(x))), & \text{if } x \neq \emptyset;\\ a, & \text{if } x = \emptyset, \end{cases}$$

by

$$\mathcal{G}' = \{\,(x,y) \mid (x \neq \emptyset \to y = \mathcal{G}(x\bigcup \mathsf{dom}(x))) \wedge (x = \emptyset \to y = a)\,\}.$$

Then there is a unique function $f\colon \omega \to V$ such that

$$\begin{aligned} f(n+1) &= \mathcal{G}'(f\restriction n+1)\\ &= \mathcal{G}((f\restriction n+1)(\underbrace{\bigcup\bigcup(n+1)}_{n}))\\ &= \mathcal{G}(f(n)),\\ f(0) &= \mathcal{G}'(\underbrace{f\restriction 0}_{\emptyset})\\ &= a. \end{aligned}$$

This concludes the proof. □

We now define

$$s_m(0) = m, \quad s_m(n+1) = s_m(n) + 1.$$

By Corollary 3.13 for every m there is such a function, and it is uniquely determined. We define

$$m + n := s_m(n).$$

Because of $s_m(1) = s_m(0+1) = s_m(0)+1 = m+1$, for $n = 1$ this definition is compatible with the previous terminology. Moreover, we have $m+0 = m$ and $m + (n+1) = (m+n)+1$.

LEMMA 3.14. (a) $m + n \in \omega$.
(b) $(m+n)+p = m+(n+p)$.
(c) $m+n = n+m$.

PROOF. (a). Induction on n. 0. Clear. $n \to n+1$. $m+(n+1) = (m+n)+1$, and by IH $m+n \in \omega$.

(b). Induction on p. 0. Clear. $p \to p+1$.

$$\begin{aligned}(m+n)+(p+1) &= [(m+n)+p]+1 && \text{by definition}\\ &= [m+(n+p)]+1 && \text{by IH}\\ &= m+[(n+p)+1]\\ &= m+[n+(p+1)].\end{aligned}$$

(c). We first prove two auxiliary propositions.

(i) $0+n = n$. The proof is by induction on n. 0. Clear. $n \to n+1$. $0+(n+1) = (0+n)+1 = n+1$.

(ii) $(m+1)+n = (m+n)+1$. Again the proof is by induction on n. 0. Clear. $n \to n+1$.

$$\begin{aligned}(m+1)+(n+1) &= [(m+1)+n]+1\\ &= [(m+n)+1]+1 && \text{by IH}\\ &= [m+(n+1)]+1.\end{aligned}$$

Now the claim $m+n = n+m$ con be proved by induction on m. 0. By (i). Step $m \to m+1$.

$$\begin{aligned}(m+1)+n &= (m+n)+1 && \text{by (ii)}\\ &= (n+m)+1 && \text{by IH}\\ &= n+(m+1).\end{aligned}$$

This concludes the proof □

We define

$$p_m(0) = 0, \quad p_m(n+1) = p_m(n) + m.$$

By Corollary 3.13 for every m there is a unique such function. Here we need

$$\mathcal{G}\colon V \to V,$$

$$\mathcal{G}(x) = \begin{cases} x+m, & \text{if } x \in \omega;\\ \emptyset, & \text{otherwise.}\end{cases}$$

We finally define $m \cdot n := p_m(n)$. Observe that this implies $m \cdot 0 = 0$, $m \cdot (n+1) = m \cdot n + m$.

LEMMA 3.15. (a) $m \cdot n \in \omega$.
(b) $m \cdot (n+p) = m \cdot n + m \cdot p$.
(c) $(n+p) \cdot m = n \cdot m + p \cdot m$.

(d) $(m \cdot n) \cdot p = m \cdot (n \cdot p)$.
(e) $0 \cdot n = 0$, $1 \cdot n = n$, $m \cdot n = n \cdot m$.

PROOF. Exercise. □

REMARK. n^m, $m - n$ can be treated similarly; later (when we deal with ordinal arithmetic) this will be done more generally. - We could now introduce the integers, rationals, reals and complex numbers in the well-known way, and prove their elementary properties.

3.4. Transitive Closure. We define the $\mathcal{R}$-transitive closure of a set a, w.r.t. a relation $\mathcal{R}$ with the property that the $\mathcal{R}$-predecessors of an arbitrary element of its domain form a set.

THEOREM 3.16. *Let $\mathcal{R}$ be a relation on $\mathcal{A}$ such that $\hat{x}^{\mathcal{R}}$ $(:= \{\, y \mid y\mathcal{R}x \,\})$ is a set, for every $x \in \mathcal{A}$. Then for every subset $a \subseteq \mathcal{A}$ there is a uniquely determined set b such that*

(a) *$a \subseteq b \subseteq \mathcal{A}$;*
(b) *b is $\mathcal{R}$-transitive;*
(c) *$\forall c. a \subseteq c \subseteq \mathcal{A} \to c$ $\mathcal{R}$-transitive $\to b \subseteq c$.*

b is called the $\mathcal{R}$-transitive closure *of a.*

PROOF. Uniqueness. Clear by (c). Existence. We shall define $f\colon \omega \to V$ by recursion on ω, such that

$$\begin{aligned} f(0) &= a, \\ f(n+1) &= \{\, y \mid \exists x \in f(n)(y\mathcal{R}x) \,\}. \end{aligned}$$

In order to apply the Recursion Theorem for ω, we must define $f(n+1)$ in the form $\mathcal{G}(f(n))$. To this end choose $\mathcal{G}\colon V \to V$, $z \mapsto \bigcup \mathsf{rng}(\mathcal{H}\restriction z)$ such that $\mathcal{H}\colon V \to V$, $x \mapsto \hat{x}$; by assumption $\mathcal{H}$ is a function. Then

$$\begin{aligned} y \in \mathcal{G}(f(n)) &\leftrightarrow y \in \bigcup \mathsf{rng}(\mathcal{H}\restriction f(n)) \\ &\leftrightarrow \exists z. z \in \mathsf{rng}(\mathcal{H}\restriction f(n)) \wedge y \in z \\ &\leftrightarrow \exists z, x. x \in f(n) \wedge z = \hat{x} \wedge y \in z \\ &\leftrightarrow \exists x. x \in f(n) \wedge y\mathcal{R}x. \end{aligned}$$

By induction on n one can see easily that $f(n)$ is a set. For 0 this is clear, and in the step $n \to n+1$ it follows - using $f(n+1) = \bigcup\{\, \hat{x} \mid x \in f(n) \,\}$ - from the IH, Replacement and the Union Axiom. – We now define $b := \bigcup \mathsf{rng}(f) = \bigcup\{\, f(n) \mid n \in \omega \,\}$. Then

(a). $a = f(0) \subseteq b \subseteq \mathcal{A}$.
(b).

$$\begin{aligned} &y\mathcal{R}x \in b \\ &y\mathcal{R}x \in f(n) \\ &y \in f(n+1) \\ &y \in b. \end{aligned}$$

(c). Let $a \subseteq c \subseteq \mathcal{A}$ and c be $\mathcal{R}$-transitive. We show $f(n) \subseteq c$ by induction on n. 0. $a \subseteq c$. $n \to n+1$.

$$y \in f(n+1)$$

$$y\mathcal{R}x \in f(n)$$
$$y\mathcal{R}x \in c$$
$$y \in c.$$

This concludes the proof. □

In the special case of the element relation $\in$ on V, the condition $\forall x(\hat{x} = \{\, y \mid y \in x \,\}$ is a set) clearly holds. Hence for every set a there is a uniquely determined $\in$-transitive closure of a. It is called the *transitive closure* of a.

By means of the notion of the $\mathcal{R}$-transitive closure we can now show that the transitively well-founded relations on $\mathcal{A}$ coincide with the well-founded relations on $\mathcal{A}$.

Let $\mathcal{R}$ be a relation on $\mathcal{A}$. $\mathcal{R}$ is a *well-founded relation* on $\mathcal{A}$ if

(a) Every nonempty subset of $\mathcal{A}$ has an $\mathcal{R}$-minimal element, i.e.,

$$\forall a.a \subseteq \mathcal{A} \to a \neq \emptyset \to \exists x{\in}a.\hat{x} \cap a = \emptyset;$$

(b) for every $x \in \mathcal{A}$, $\hat{x}$ is a set.

THEOREM 3.17. *The transitively well-founded relations on $\mathcal{A}$ are the same as the well-founded relations on $\mathcal{A}$.*

PROOF. Every transitively well-founded relation on $\mathcal{A}$ is well-founded by Lemma 3.1(b). Conversely, every well-founded relation on $\mathcal{A}$ is transitively well-founded, since for every $x \in \mathcal{A}$, the $\mathcal{R}$-transitive closure of $\hat{x}$ is an $\mathcal{R}$-transitive $b \subseteq \mathcal{A}$ such that $\hat{x} \subseteq b$. □

Therefore, the Induction Theorem 3.2 and the Recursion Theorem 3.3 also hold for well-founded relations. Moreover, by Lemma 3.1(a), every nonempty sub*class* of a well-founded relation $\mathcal{R}$ has an $\mathcal{R}$-minimal element.

Later we will require the so-called Regularity Axiom, which says that the relation $\in$ on V is well-founded, i.e.,

$$\forall a.a \neq \emptyset \to \exists x{\in}a.x \cap a = \emptyset.$$

This will provide us with an important example of a well-founded relation.

We now consider extensional well-founded relations. From the Regularity Axiom it will follow that the $\in$-relation on an arbitrary class $\mathcal{A}$ is a well-founded extensional relation. Here we show - even without the Regularity Axiom - the converse, namely that every well-founded extensional relation is isomorphic to the $\in$-relation on a transitive class. This is Mostowski's Isomorphy Theorem. Then we consider linear well-founded orderings, well-orderings for short. They are always extensional, and hence isomorphic to the $\in$-relation on certain classes, which will be called ordinal classes. Ordinals will then be defined as ordinal sets.

A relation $\mathcal{R}$ on $\mathcal{A}$ is *extensional* if for all $x, y \in \mathcal{A}$

$$(\forall z{\in}\mathcal{A}.z\mathcal{R}x \leftrightarrow z\mathcal{R}y) \to x = y.$$

For example, for a transitive class $\mathcal{A}$ the relation $\in \cap(\mathcal{A} \times \mathcal{A})$ is extensional on $\mathcal{A}$. This can be seen as follows. Let $x, y \in \mathcal{A}$. For $\mathcal{R} := \in \cap(\mathcal{A} \times \mathcal{A})$ we have $z\mathcal{R}x \leftrightarrow z \in x$, since $\mathcal{A}$ is transitive. We obtain

$$\forall z{\in}\mathcal{A}.z\mathcal{R}x \leftrightarrow z\mathcal{R}y$$
$$\forall z.z \in x \leftrightarrow z \in y$$

$$x = y$$

From the Regularity Axiom it will follows that all these relations are well-founded. But even without the Regularity Axiom these relations have a distinguished meaning; cf. Corollary 3.19.

THEOREM 3.18 (Isomorphy Theorem of Mostowski). *Let $\mathcal{R}$ be a well-founded extensional relation on $\mathcal{A}$. Then there is a unique isomorphism $\mathcal{F}$ of $\mathcal{A}$ onto a transitive class $\mathcal{B}$, i.e.*

$$\exists^{=1}\mathcal{F}.\mathcal{F}\colon \mathcal{A} \leftrightarrow \mathsf{rng}(\mathcal{F}) \wedge \mathsf{rng}(\mathcal{F}) \text{ transitive } \wedge \forall x, y{\in}\mathcal{A}.y\mathcal{R}x \leftrightarrow \mathcal{F}(y) \in \mathcal{F}(x).$$

PROOF. Existence. We define by the Recursion Theorem

$$\begin{aligned}&\mathcal{F}\colon \mathcal{A} \to V,\\ &\mathcal{F}(x) = \mathsf{rng}(\mathcal{F}{\restriction}\hat{x}) \quad (= \{\, \mathcal{F}(y) \mid y\mathcal{R}x \,\}).\end{aligned}$$

$\mathcal{F}$ injective: We show $\forall x, y{\in}\mathcal{A}.\mathcal{F}(x) = \mathcal{F}(y) \to x = y$ by $\mathcal{R}$-induction on x. So let $x, y \in \mathcal{A}$ be given such that $\mathcal{F}(x) = \mathcal{F}(y)$. By IH

$$\forall z{\in}\mathcal{A}.z\mathcal{R}x \to \forall u{\in}\mathcal{A}.\mathcal{F}(z) = \mathcal{F}(u) \to z = u.$$

It suffices to show $z\mathcal{R}x \leftrightarrow z\mathcal{R}y$, for all $z \in \mathcal{A}$. $\to$.

$$\begin{aligned}&z\mathcal{R}x\\ &\mathcal{F}(z) \in \mathcal{F}(x) = \mathcal{F}(y) = \{\, \mathcal{F}(u) \mid u\mathcal{R}y \,\}\\ &\mathcal{F}(z) = \mathcal{F}(u) && \text{for some } u\mathcal{R}y\\ &z = u && \text{by IH, since } z\mathcal{R}x\\ &z\mathcal{R}y\end{aligned}$$

$\leftarrow$.

$$\begin{aligned}&z\mathcal{R}y\\ &\mathcal{F}(z) \in \mathcal{F}(y) = \mathcal{F}(x) = \{\, \mathcal{F}(u) \mid u\mathcal{R}x \,\}\\ &\mathcal{F}(z) = \mathcal{F}(u) && \text{for some } u\mathcal{R}x\\ &z = u && \text{by IH, since } u\mathcal{R}x\\ &z\mathcal{R}x.\end{aligned}$$

$\mathsf{rng}(\mathcal{F})$ is transitive: Assume $u \in v \in \mathsf{rng}(\mathcal{F})$. Then $v = \mathcal{F}(x)$ for some $x \in \mathcal{A}$, hence $u = \mathcal{F}(y)$ for some $y\mathcal{R}x$.

$y\mathcal{R}x \leftrightarrow \mathcal{F}(y) \in \mathcal{F}(x)$: $\to$. Assume $y\mathcal{R}x$. Then $\mathcal{F}(y) \in \mathcal{F}(x)$ by definition of $\mathcal{F}$. $\leftarrow$.

$$\begin{aligned}&\mathcal{F}(y) \in \mathcal{F}(x) = \{\, \mathcal{F}(z) \mid z\mathcal{R}x \,\}\\ &\mathcal{F}(y) = \mathcal{F}(z) && \text{for some } z\mathcal{R}x\\ &y = z && \text{since } \mathcal{F} \text{ is injective}\\ &y\mathcal{R}x.\end{aligned}$$

Uniqueness. Let $\mathcal{F}_i$ $(i = 1, 2)$ be two isomorphisms as described in the theorem. We show $\forall x{\in}\mathcal{A}\,(\mathcal{F}_1(x) = \mathcal{F}_2(x))$, by $\mathcal{R}$-induction on x. By symmetry it suffices to prove $u \in \mathcal{F}_1(x) \to u \in \mathcal{F}_2(x)$.

$$\begin{aligned}&u \in \mathcal{F}_1(x)\\ &u = \mathcal{F}_1(y) && \text{for some } y \in \mathcal{A}, \text{ since } \mathsf{rng}(\mathcal{F}_1) \text{ is transitive}\end{aligned}$$

$y\mathcal{R}x$	by the isomorphy condition for $\mathcal{F}_1$
$u = \mathcal{F}_2(y)$	by IH
$\mathcal{F}_2(y) \in \mathcal{F}_2(x)$	by the isomorphy condition for $\mathcal{F}_2$
$u \in \mathcal{F}_2(x)$.	

This concludes the proof. □

A relation $\mathcal{R}$ on $\mathcal{A}$ is a *linear ordering* if for all $x, y, z \in \mathcal{A}$

$\neg x\mathcal{R}x$	irreflexivity,
$x\mathcal{R}y \to y\mathcal{R}z \to x\mathcal{R}z$	transitivity,
$x\mathcal{R}y \vee x = y \vee y\mathcal{R}x$	trichotomy (or compatibility).

$\mathcal{R}$ is a *well-ordering* if $\mathcal{R}$ is a well-founded linear ordering.

REMARK. Every well-ordering $\mathcal{R}$ on $\mathcal{A}$ is extensional. To see this, assume

$$\forall z{\in}\mathcal{A}.z\mathcal{R}x \leftrightarrow z\mathcal{R}y.$$

Then $x = y$ by trichotomy, since from $x\mathcal{R}y$ we obtain by assumption $x\mathcal{R}x$, contradicting irreflexivity, and similarly $y\mathcal{R}x$ entails a contradiction.

COROLLARY 3.19. *For every well-ordering $\mathcal{R}$ on $\mathcal{A}$ there is a unique isomorphism $\mathcal{F}$ of $\mathcal{A}$ onto a transitive class $\mathcal{B}$.*

3.5. Ordinal classes and ordinals. We now study more closely the transitive classes that appear as images of well-orderings.

$\mathcal{A}$ is an *ordinal class* if $\mathcal{A}$ is transitive and $\in \cap(\mathcal{A} \times \mathcal{A})$ is a well-ordering on $\mathcal{A}$. Ordinal classes that happen to be sets are called *ordinals*. Define

$$\mathsf{On} := \{\, x \mid x \text{ is an ordinal} \,\}.$$

First we give a convenient characterization of ordinal classes. $\mathcal{A}$ is called *connex* if for all $x, y \in \mathcal{A}$

$$x \in y \vee x = y \vee y \in x.$$

For instance, ω is connex by Lemma 3.7(d). Also every n is connex, since by Lemma 3.5(b), ω is transitive.

A class $\mathcal{A}$ is *well-founded* if $\in \cap(\mathcal{A} \times \mathcal{A})$ is a well-founded relation on $\mathcal{A}$, i.e., if

$$\forall a.a \subseteq \mathcal{A} \to a \neq \emptyset \to \exists x{\in}a.x \cap a = \emptyset.$$

We now show that in well-founded classes there can be no finite $\in$-cycles.

LEMMA 3.20. *Let $\mathcal{A}$ be well-founded. Then for arbitrary $x_1, \ldots, x_n \in \mathcal{A}$ we can never have*

$$x_1 \in x_2 \in \cdots \in x_n \in x_1.$$

PROOF. Assume $x_1 \in x_2 \in \cdots \in x_n \in x_1$. Consider $\{x_1, \ldots, x_n\}$. Since $\mathcal{A}$ is well-founded we may assume $x_1 \cap \{x_1, \ldots, x_n\} = \emptyset$. But this contradicts $x_n \in x_1$. □

COROLLARY 3.21. *$\mathcal{A}$ is an ordinal class iff $\mathcal{A}$ is transitive, connex and well-founded.*

PROOF. $\rightarrow$ is clear; $\mathcal{A}$ is connex because of trichotomy. $\leftarrow$: We must show, for all $x, y, z \in \mathcal{A}$,

$$x \notin x,$$
$$x \in y \rightarrow y \in z \rightarrow x \in z.$$

Since $\mathcal{A}$ is connex, both propositions follow from Lemma 3.20. $\square$

Here are some examples of ordinals. ω is transitive by Lemma 3.5(b), connex as noted above and well-founded by the principle of least element (Theorem 3.10). So, ω is an ordinal class. Since ω by the Infinity Axiom is a set, ω is even an ordinal. Also, n is transitive by Lemma 3.5(a), connex (see above) and well-founded; the latter follows with transitivity of ω by the principle of least element (Theorem 3.10).

Let us write $\mathsf{Ord}(\mathcal{A})$ for "$\mathcal{A}$ is an ordinal class". We now show that ordinal classes have properties similar to those of natural numbers: the relation $\in$ has the properties of $<$, and the relation $\subseteq$ has the properties of $\leq$.

LEMMA 3.22. (a) $\mathsf{Ord}(\mathcal{A}) \rightarrow \mathsf{Ord}(\mathcal{B}) \rightarrow \mathsf{Ord}(\mathcal{A} \cap \mathcal{B})$.
(b) $\mathsf{Ord}(\mathcal{A}) \rightarrow x \in \mathcal{A} \rightarrow \mathsf{Ord}(x)$.
(c) $\mathsf{Ord}(\mathcal{A}) \rightarrow \mathsf{Ord}(\mathcal{B}) \rightarrow (\mathcal{A} \subseteq \mathcal{B} \leftrightarrow \mathcal{A} \in \mathcal{B} \vee \mathcal{A} = \mathcal{B})$.
(d) $\mathsf{Ord}(\mathcal{A}) \rightarrow \mathsf{Ord}(\mathcal{B}) \rightarrow (\mathcal{A} \in \mathcal{B} \vee \mathcal{A} = \mathcal{B} \vee \mathcal{B} \in \mathcal{A})$.

PROOF. (a). $\mathcal{A} \cap \mathcal{B}$ transitive:

$$x \in y \in \mathcal{A} \cap \mathcal{B}$$
$$x \in y \in \mathcal{A} \quad \text{and} \quad x \in y \in \mathcal{B}$$
$$x \in \mathcal{A} \quad \text{and} \quad x \in \mathcal{B}$$
$$x \in \mathcal{A} \cap \mathcal{B}.$$

$\mathcal{A} \cap \mathcal{B}$ connex, well-founded: Clear.

(b). x transitive:

$$u \in v \in x \in \mathcal{A}$$
$$u \in v \in \mathcal{A}$$
$$u \in \mathcal{A}$$
$$u \in x \vee u = x \vee x \in u.$$

From $u = x$ it follows that $u \in v \in u$ contradicting Lemma 3.20, and from $x \in u$ it follows that $u \in v \in x \in u$, again contradicting Lemma 3.20.

x connex, well-founded. Clear, for $x \subseteq \mathcal{A}$.

(c). $\leftarrow$. Clear, for $\mathcal{B}$ is transitive. $\rightarrow$. Let $\mathcal{A} \subseteq \mathcal{B}$. Wlog $\mathcal{A} \subsetneq \mathcal{B}$. Choose $x \in \mathcal{B} \setminus \mathcal{A}$ such that $x \cap (\mathcal{B} \setminus \mathcal{A}) = \emptyset$ (this is possible, since $\mathcal{B}$ is well-founded); it suffices to show that $x = \mathcal{A}$.

$x \subseteq \mathcal{A}$. Assume $y \in x$, hence $y \in x \in \mathcal{B}$. Then $y \in \mathcal{A}$, for $x \cap (\mathcal{B} \setminus \mathcal{A}) = \emptyset$.

$\mathcal{A} \subseteq x$. Assume $y \in \mathcal{A}$. Then also $y \in \mathcal{B}$. It follows that $x \in y \vee x = y \vee y \in x$. But the first two cases are impossible, for in both of them we obtain $x \in \mathcal{A}$.

(d). Assume $\mathsf{Ord}(\mathcal{A})$ and $\mathsf{Ord}(\mathcal{B})$. Then by (a), $\mathsf{Ord}(\mathcal{A} \cap \mathcal{B})$. Using (c) we obtain

$$[(\mathcal{A} \cap \mathcal{B} \in \mathcal{A}) \vee (\mathcal{A} \cap \mathcal{B} = \mathcal{A})] \wedge [(\mathcal{A} \cap \mathcal{B} \in \mathcal{B}) \vee (\mathcal{A} \cap \mathcal{B} = \mathcal{B})].$$

Distributing yields

$$(\mathcal{A} \cap \mathcal{B} \in \mathcal{A} \cap \mathcal{B}) \vee (\mathcal{A} \in \mathcal{B}) \vee (\mathcal{B} \in \mathcal{A}) \vee (\mathcal{A} = \mathcal{B}).$$

But the first case $\mathcal{A} \cap \mathcal{B} \in \mathcal{A} \cap \mathcal{B}$ is impossible by Lemma 3.20. □

LEMMA 3.23. (a) $\mathsf{Ord}(\mathsf{On})$.
(b) On *is not a set.*
(c) On *is the only proper ordinal class.*

PROOF. (a). On is transitive by Lemma 3.22(b) and it is connex by Lemma 3.22(d). On is well-founded: Let $a \subseteq \mathsf{On}$, $a \neq \emptyset$. Choose $x \in a$. Wlog $x \cap a \neq \emptyset$. Since x is well-founded, there is a $y \in x \cap a$ such that $y \cap x \cap a = \emptyset$. It follows that $y \in a$ and $y \cap a = \emptyset$; the latter holds since $y \subseteq x$ because of $y \in x$, x transitive.

(b). Assume On is a set. Then $\mathsf{On} \in \mathsf{On}$, contradicting Lemma 3.20.

(c). Let $\mathsf{Ord}(\mathcal{A})$, $\mathcal{A}$ not a set. By Lemma 3.22(d)

$$\mathcal{A} \in \mathsf{On} \vee \mathcal{A} = \mathsf{On} \vee \mathsf{On} \in \mathcal{A}.$$

The first and the last case are excluded, for then $\mathcal{A}$ (or On, resp.) would be a set. □

LEMMA 3.24. (a) On *is inductive,*
(b) $n, \omega \in \mathsf{On}$.

PROOF. (a). $0 \in \mathsf{On}$ is clear. So let $x \in \mathsf{On}$. We must show $x + 1 \in \mathsf{On}$, that is $x \cup \{x\} \in \mathsf{On}$.

$x \cup \{x\}$ transitive: Assume $u \in v \in x \cup \{x\}$, so $u \in v \in x$ or $u \in v = x$. In both cases it follows that $u \in x$.

$x \cup \{x\}$ is connex: Assume $u, v \in x \cup \{x\}$. Then

$$u, v \in x \vee (u \in x \wedge v = x) \vee (u = x \wedge v \in x) \vee (u = v = x)$$
$$u \in v \vee u = v \vee v \in u.$$

$x \cup \{x\}$ is well-founded: Let $a \subseteq x \cup \{x\}$, $a \neq \emptyset$. We must show $\exists y \in a\, (y \cap a = \emptyset)$. **Case** $a \cap x \neq \emptyset$. Then the claim follows from the well-foundedness of x. **Case** $a \cap x = \emptyset$. Then $a = \{x\}$, and we have $x \cap \{x\} = \emptyset$.

(b). This has been proved above, after Corollary 3.21. □

LEMMA 3.25. $x, y \in \mathsf{On} \to x + 1 = y + 1 \to x = y$.

PROOF. The proof is similar to the proof of the second Peano-Axiom in Theorem 3.8(b).

$$x + 1 = y + 1$$
$$x \in y + 1 \wedge y \in x + 1$$
$$(x \in y \wedge y \in x) \vee x = y.$$

Since the first case is impossible by Lemma 3.22, we have $x = y$. □

LEMMA 3.26. $\mathcal{A} \subseteq \mathsf{On} \to \bigcup \mathcal{A} \in \mathsf{On} \vee \bigcup \mathcal{A} = \mathsf{On}$.

PROOF. It suffices to show $\mathsf{Ord}(\bigcup \mathcal{A})$. $\bigcup \mathcal{A}$ is transitive: Let $x \in y \in \bigcup \mathcal{A}$, so $x \in y \in z \in \mathcal{A}$ for some z. Then we have $x \in z \in \mathcal{A}$, since $\mathcal{A} \subseteq \mathsf{On}$. Hence $x \in \bigcup \mathcal{A}$.

$\bigcup \mathcal{A}$ is connex and well-founded: It suffices to prove $\bigcup \mathcal{A} \subseteq \mathsf{On}$. So let $x \in \bigcup \mathcal{A}$, hence $x \in y \in \mathcal{A}$ for some y. Then $x \in y$ and $y \in \mathsf{On}$, so $x \in \mathsf{On}$. □

REMARK. If $\mathcal{A} \subseteq \mathsf{On}$, then $\bigcup \mathcal{A}$ is the least upper bound of $\mathcal{A}$ w.r.t. the well-ordering $\in \cap(\mathsf{On} \times \mathsf{On})$ of On, for by definition of $\bigcup \mathcal{A}$ we have

$$x \in \mathcal{A} \to x \subseteq \bigcup \mathcal{A},$$
$$(\forall x \in \mathcal{A}. x \subseteq y) \to \bigcup \mathcal{A} \subseteq y.$$

We therefore also write $\sup \mathcal{A}$ for $\bigcup \mathcal{A}$.

Here are some examples of ordinals:

$$\begin{aligned}
&0\\
&1 = 0 + 1\\
&2 = 1 + 1\\
&\vdots\\
&\omega \quad \text{set by the Infinity Axiom}\\
&\omega + 1\\
&\omega + 2\\
&\vdots\\
&\omega \cdot 2 := \bigcup\{\, \omega + n \mid n \in \omega \,\} \qquad \text{by recursion on } \omega\\
&\omega \cdot 2 + 1\\
&\omega \cdot 2 + 2\\
&\vdots\\
&\omega \cdot 3 := \bigcup\{\, \omega \cdot 2 + n \mid n \in \omega \,\}\\
&\vdots\\
&\omega \cdot 4\\
&\vdots\\
&\omega \cdot \omega := \omega^2 := \bigcup\{\, \omega \cdot n \mid n \in \omega \,\}\\
&\omega^2 + 1\\
&\omega^2 + 2\\
&\vdots\\
&\omega^2 + \omega\\
&\omega^2 + \omega + 1\\
&\omega^2 + \omega + 2
\end{aligned}$$

$$\vdots$$
$$\omega^2 + \omega \cdot 2$$
$$\vdots$$
$$\omega^2 + \omega \cdot 3$$
$$\vdots$$
$$\omega^3$$
$$\vdots$$
$$\omega^4$$
$$\vdots$$
$$\omega^\omega$$
$$\omega^\omega + 1$$
$$\vdots$$

and so on.

α, β, γ will denote ordinals.

α is a *successor number* if $\exists\beta(\alpha = \beta + 1)$. α is a *limit* if α is neither 0 nor a successor number. We write

$$\mathsf{Lim}(\alpha) \text{ for } \alpha \neq 0 \wedge \neg\exists\beta(\alpha = \beta + 1).$$

Clearly for arbitrary α either $\alpha = 0$ or α is successor number or α is a limit.

LEMMA 3.27. (a) $\mathsf{Lim}(\alpha) \leftrightarrow \alpha \neq 0 \wedge \forall\beta.\beta \in \alpha \to \beta + 1 \in \alpha$.
(b) $\mathsf{Lim}(\omega)$.
(c) $\mathsf{Lim}(\alpha) \to \omega \subseteq \alpha$.

PROOF. (a). $\to$: Let $\beta \in \alpha$. Then $\beta + 1 \in \alpha \vee \beta + 1 = \alpha \vee \alpha \in \beta + 1$. The second case $\beta + 1 = \alpha$ is excluded assumption. In the third case it follows that $\alpha \in \beta \vee \alpha = \beta$; but because of $\beta \in \alpha$ both are impossible by Lemma 3.20. $\leftarrow$. Let $\alpha \neq 0$ and assume $\forall\beta.\beta \in \alpha \to \beta + 1 \in \alpha$. Then if α is not a limit, we must have $\alpha = \beta + 1$. Then we obtain $\beta \in \alpha$, hence by assumption also $\beta + 1 \in \alpha$ and hence $\alpha \in \alpha$, which is impossible.

(b). Follows from (a), since ω is inductive.

(c). Assume $\mathsf{Lim}(\alpha)$. We show $n \in \alpha$ by induction on n. 0. We have $0 \in \alpha \vee 0 = \alpha \vee \alpha \in 0$, where the cases two and three clearly are impossible. $n + 1$. We have $n \in \alpha$ by IH, hence $n + 1 \in \alpha$ by (a). □

LEMMA 3.28. (a) $\alpha = \bigcup_{\beta\in\alpha}(\beta + 1)$.
(b) *For limits* α *we have* $\alpha = \bigcup_{\beta\in\alpha}\beta$.

PROOF. (a). $\subseteq$. Let $\beta \in \alpha$. The claim follows from $\beta \in \beta + 1$. $\supseteq$. Let $\beta \in \alpha$. Then $\beta + 1 \subseteq \alpha$.

(b). $\subseteq$. Let $\gamma \in \alpha$. Then $\gamma \in \gamma + 1 \in \alpha$. $\supseteq$. Let $\gamma \in \beta \in \alpha$. We obtain $\gamma \in \alpha$. □

THEOREM 3.29 (Transfinite induction on On; class form).

$$(\forall\alpha.\alpha \subseteq \mathcal{B} \to \alpha \in \mathcal{B}) \to \mathsf{On} \subseteq \mathcal{B}.$$

PROOF. This is a special case of the Induction Theorem 3.2 □

COROLLARY 3.30 (Different forms of transfinite induction on On). *First form:*

$$\begin{aligned} A(0) &\to (\forall\alpha.A(\alpha) \to A(\alpha+1)) \\ &\to (\forall\alpha.\mathsf{Lim}(\alpha) \to (\forall\beta.\beta \in \alpha \to A(\beta)) \to A(\alpha)) \\ &\to \forall\alpha A(\alpha). \end{aligned}$$

Second form: (Transfinite induction on On*, using all predecessors).*

$$[\forall\alpha.(\forall\beta.\beta \in \alpha \to A(\beta)) \to A(\alpha)] \to \forall\alpha A(\alpha).$$

Third form: (Principle of least element for On).

$$\exists\alpha A(\alpha) \to \exists\alpha.A(\alpha) \wedge \neg\exists\beta.\beta \in \alpha \wedge A(\beta).$$

PROOF. The third form follows from the second by contraposition. Also, the first form follows easily from the second. The second form follows from Theorem 3.29 using $\mathcal{B} := \{\, \alpha \mid A(\alpha) \,\}$. □

THEOREM 3.31 (Transfinite recursion on On). *Let $\mathcal{G}\colon V \to V$. Then there is exactly one function $\mathcal{F}\colon \mathsf{On} \to V$ such that for all α*

$$\mathcal{F}(\alpha) = \mathcal{G}(\mathcal{F}{\restriction}\alpha).$$

PROOF. This is a special case of the Recursion Theorem 3.3. □

COROLLARY 3.32. *Assume $\mathcal{G}\colon V \to V$, $\mathcal{H}\colon V \to V$ and a is a set. Then there is a unique function $\mathcal{F}\colon \mathsf{On} \to V$ such that*

$$\begin{aligned} &\mathcal{F}(0) = a, \\ &\mathcal{F}(\alpha+1) = \mathcal{G}(\mathcal{F}(\alpha)), \\ &\mathcal{F}(\alpha) = \mathcal{H}(\mathcal{F}{\restriction}\alpha) \qquad \textit{for } \alpha \textit{ limit.} \end{aligned}$$

PROOF. First observe that $\bigcup(\alpha+1) = \alpha$, because of

$$\begin{aligned} \gamma \in \bigcup(\alpha+1) &\leftrightarrow \exists\beta.\gamma \in \beta \in \alpha+1 \\ &\leftrightarrow \exists\beta.\gamma \in \beta \subseteq \alpha \\ &\leftrightarrow \gamma \in \alpha. \end{aligned}$$

For given a, $\mathcal{G}$ and $\mathcal{H}$ we shall find a $\mathcal{G}'$ such that

$$\begin{aligned} &\mathcal{G}'(0) = a, \\ &\mathcal{G}'(\mathcal{F}{\restriction}\alpha+1) = \mathcal{G}(\mathcal{F}(\alpha)), \\ &\mathcal{G}'(\mathcal{F}{\restriction}\alpha) = \mathcal{H}(\mathcal{F}{\restriction}\alpha) \qquad \text{for } \alpha \text{ limit.} \end{aligned}$$

We define a function $\mathcal{G}'\colon V \to V$ by

$$\mathcal{G}'(x) = \begin{cases} a, & \text{otherwise;} \\ \mathcal{G}(x(\bigcup \mathsf{dom}(x))), & \text{if } \exists\beta(\mathsf{dom}(x) = \beta+1); \\ \mathcal{H}(x), & \text{if } \mathsf{Lim}(\mathsf{dom}(x)). \end{cases}$$

By the Recursion Theorem 3.3 there is a unique $\mathcal{F}\colon \mathsf{On} \to V$ such that, for all α,

$$\mathcal{F}(\alpha) = \mathcal{G}'(\mathcal{F}\restriction\alpha).$$

Clearly this property of $\mathcal{F}$ is equivalent to the equations above. □

3.6. Regularity Axiom, Von Neumann Levels, Rank. Recall the cumulative type structure:

Level 0: $-$
Level 1: $\emptyset$
Level 2: $\emptyset, \{\emptyset\}$
Level 3: $\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}$
and so on.

Using ordinals we can now consider transfinite levels as well. The level ω consists of all sets whose elements are formed on finite levels, and the level $\omega+1$ consists of all sets whose elements are formed on finite levels or at level ω, and so on. Generally we define the *Von Neumann levels* V_α as follows, by transfinite recursion on On.

$$\begin{aligned} &V_0 = \emptyset, \\ &V_{\alpha+1} = \mathcal{P}(V_\alpha), \\ &V_\alpha = \bigcup_{\beta\in\alpha} V_\beta \qquad \text{for } \alpha \text{ limit.} \end{aligned}$$

Remark. More precisely, $V_\alpha := \mathcal{F}(\alpha)$, where $\mathcal{F}\colon \mathsf{On} \to V$ is defined as follows, by transfinite recursion on On:

$$\begin{aligned} &\mathcal{F}(0) = \emptyset, \\ &\mathcal{F}(\alpha+1) = \mathcal{P}(\mathcal{F}(\alpha)), \\ &\mathcal{F}(\alpha) = \bigcup \mathsf{rng}(\mathcal{F}\restriction\alpha) \qquad \text{for } \alpha \text{ limit.} \end{aligned}$$

Lemma 3.33. (a) *V_α is transitive.*
(b) *$\alpha \in \beta \to V_\alpha \in V_\beta$.*
(c) *$\alpha \subseteq \beta \to V_\alpha \subseteq V_\beta$.*
(d) *$V_\alpha \cap \mathsf{On} = \alpha$.*

Proof. (a). (Transfinite) induction by α. 0. $\emptyset$ is transitive. $\alpha+1$.

$$\begin{aligned} &x \in y \in V_{\alpha+1} = \mathcal{P}(V_\alpha) \\ &x \in y \subseteq V_\alpha \\ &x \in V_\alpha \\ &x \subseteq V_\alpha && \text{by IH} \\ &x \in V_{\alpha+1}. \end{aligned}$$

α limit.

$$\begin{aligned} &x \in y \in V_\alpha = \bigcup_{\beta\in\alpha} V_\beta \\ &x \in y \in V_\beta && \text{for some } \beta \in \alpha \\ &x \in V_\beta && \text{by IH} \end{aligned}$$

$$x \in V_\alpha.$$

(b). Induction by β. 0. Clear. $\beta + 1$.

$$\begin{aligned}
&\alpha \in \beta + 1 \\
&\alpha \in \beta \text{ or } \alpha = \beta \\
&V_\alpha \in V_\beta \text{ or } V_\alpha = V_\beta && \text{by IH} \\
&V_\alpha \subseteq V_\beta && \text{by (a)} \\
&V_\alpha \in V_{\beta+1}.
\end{aligned}$$

β limit.

$$\begin{aligned}
&\alpha \in \beta \\
&\alpha + 1 \in \beta \\
&V_\alpha \in V_{\alpha+1} \subseteq \bigcup_{\gamma\in\beta} V_\gamma = V_\beta.
\end{aligned}$$

(c). Using $\alpha \subseteq \beta \leftrightarrow \alpha \in \beta \vee \alpha = \beta$ the claim follows from (a) and (b).

(d). Induction on α. 0. Clear. $\alpha + 1$.

$$\begin{aligned}
\beta \in V_{\alpha+1} &\leftrightarrow \beta \subseteq V_\alpha \\
&\leftrightarrow \beta \subseteq V_\alpha \cap \mathsf{On} = \alpha && \text{by IH} \\
&\leftrightarrow \beta \in \alpha + 1.
\end{aligned}$$

α limit.

$$\begin{aligned}
V_\alpha \cap \mathsf{On} &= (\bigcup_{\beta\in\alpha} V_\beta) \cap \mathsf{On} \\
&= \bigcup_{\beta\in\alpha} (V_\beta \cap \mathsf{On}) \\
&= \bigcup_{\beta\in\alpha} \beta && \text{by IH} \\
&= \alpha.
\end{aligned}$$

This concludes the proof. □

We now show that the von Neumann levels exhaust the universe, which means that $V = \bigcup_{\alpha\in\mathsf{On}} V_\alpha$. However, this requires another axiom, the *Regularity Axiom*, which says that the relation $\in$ on V is well-founded, i.e.,

Axiom (Regularity Axiom).

$$\forall a.a \neq \emptyset \to \exists x{\in}a\,(x \cap a = \emptyset).$$

We want to assign to every set x an ordinal α, namely the least α such that $x \subseteq V_\alpha$. To this end we need the notion of the *rank* $\mathsf{rn}(x)$ of a set x, which is defined recursively by

$$\mathsf{rn}(x) := \bigcup\{\, \mathsf{rn}(y) + 1 \mid y \in x \,\}.$$

More precisely we define $\mathsf{rn}(x) := \mathcal{F}(x)$, where $\mathcal{F}\colon V \to V$ is defined as follows (using the Recursion Theorem 3.3 for well-founded relations):

$$\mathcal{F}(x) := \bigcup \mathsf{rng}(\mathcal{H}(\mathcal{F}{\restriction}x))$$

mit

$$\mathcal{H}(z) := \{\, (u, v+1) \mid (u,v) \in z \,\}.$$

We first show that $\mathsf{rn}(x)$ has the property formulated above.

LEMMA 3.34. (a) $\mathsf{rn}(x) \in \mathsf{On}$.
(b) $x \subseteq V_{\mathsf{rn}(x)}$.
(c) $x \subseteq V_\alpha \to \mathsf{rn}(x) \subseteq \alpha$.

PROOF. (a). $\in$-induction on x. We have $\mathsf{rn}(x) = \bigcup\{\, \mathsf{rn}(y)+1 \mid y \in x \,\} \in \mathsf{On}$, for by IH $\mathsf{rn}(y) \in \mathsf{On}$ for every $y \in x$.

(b). $\in$-induction on x. Let $y \in x$. Then $y \subseteq V_{\mathsf{rn}(y)}$ by IH, hence $y \in \mathcal{P}(V_{\mathsf{rn}(y)}) = V_{\mathsf{rn}(y)+1} \subseteq V_{\mathsf{rn}(x)}$ because of $\mathsf{rn}(y)+1 \subseteq \mathsf{rn}(x)$.

(c). Induction on α. Let $x \subseteq V_\alpha$. We must show $\mathsf{rn}(x) = \bigcup\{\, \mathsf{rn}(y)+1 \mid y \in x \,\} \subseteq \alpha$. Let $y \in x$. We must show $\mathsf{rn}(y)+1 \subseteq \alpha$. Because of $x \subseteq V_\alpha$ we have $y \in V_\alpha$. This implies $y \subseteq V_\beta$ for some $\beta \in \alpha$, for in case $\alpha = \alpha' + 1$ we have $y \in V_{\alpha'+1} = \mathcal{P}(V_{\alpha'})$ and hence $y \subseteq V_{\alpha'}$, and in case α limit we have $y \in V_\alpha = \bigcup_{\beta\in\alpha} V_\beta$, hence $y \in V_\beta$ and therefore $y \subseteq V_\beta$ for some $\beta \in \alpha$. - By IH it follows that $\mathsf{rn}(y) \subseteq \beta$, whence $\mathsf{rn}(y) \in \alpha$. □

Now we obtain easily the proposition formulated above as our goal.

COROLLARY 3.35. $V = \bigcup_{\alpha\in\mathsf{On}} V_\alpha$.

PROOF. $\supseteq$ is clear. $\subseteq$. For every x we have $x \subseteq V_{\mathsf{rn}(x)}$ by Lemma 3.34(b), hence $x \in V_{\mathsf{rn}(x)+1}$. □

Now V_α can be characterized as the set of all sets of rank less than α.

LEMMA 3.36. $V_\alpha = \{\, x \mid \mathsf{rn}(x) \in \alpha \,\}$.

PROOF. $\supseteq$. Let $\mathsf{rn}(x) \in \alpha$. Then $x \subseteq V_{\mathsf{rn}(x)}$ implies $x \in V_{\mathsf{rn}(x)+1} \subseteq V_\alpha$.

$\subseteq$. Induction on α. **Case** 0. Clear. **Case** $\alpha+1$. Let $x \in V_{\alpha+1}$. Then $x \in \mathcal{P}(V_\alpha)$, hence $x \subseteq V_\alpha$. For every $y \in x$ we have $y \in V_\alpha$ and hence $\mathsf{rn}(y) \in \alpha$ by IH, so $\mathsf{rn}(y)+1 \subseteq \alpha$. Therefore $\mathsf{rn}(x) = \bigcup\{\, \mathsf{rn}(y)+1 \mid y \in x \,\} \subseteq \alpha$. **Case** α limit. Let $x \in V_\alpha$. Then $x \in V_\beta$ for some $\beta \in \alpha$, hence $\mathsf{rn}(x) \in \beta$ by IH, hence $\mathsf{rn}(x) \in \alpha$. □

From $x \in y$ and $x \subseteq y$, resp., we can infer the corresponding relations between the ranks.

LEMMA 3.37. (a) $x \in y \to \mathsf{rn}(x) \in \mathsf{rn}(y)$.
(b) $x \subseteq y \to \mathsf{rn}(x) \subseteq \mathsf{rn}(y)$.

PROOF. (a). Because of $\mathsf{rn}(y) = \bigcup\{\, \mathsf{rn}(x)+1 \mid x \in y \,\}$ this is clear. (b). For every $z \in x$ we have $\mathsf{rn}(z) \in \mathsf{rn}(y)$ by (a), hence $\mathsf{rn}(x) = \bigcup\{\, \mathsf{rn}(z)+1 \mid z \in x \,\} \subseteq \mathsf{rn}(y)$. □

Moreover we can show that the sets α and V_α both have rank α.

LEMMA 3.38. (a) $\mathsf{rn}(\alpha) = \alpha$.
(b) $\mathsf{rn}(V_\alpha) = \alpha$.

PROOF. (a). Induction on α. We have $\mathsf{rn}(\alpha) = \bigcup\{\, \mathsf{rn}(\beta)+1 \mid \beta \in \alpha \,\}$, hence by IH $\mathsf{rn}(\alpha) = \bigcup\{\, \beta+1 \mid \beta \in \alpha \,\} = \alpha$ by Lemma 3.28(a).

(b). We have

$$\mathsf{rn}(V_\alpha) = \bigcup\{\, \mathsf{rn}(x)+1 \mid x \in V_\alpha \,\}$$

$$= \bigcup \{ \mathsf{rn}(x) + 1 \mid \mathsf{rn}(x) \in \alpha \} \quad \text{by Lemma 3.36}$$
$$\subseteq \alpha.$$

Conversely, let $\beta \in \alpha$. By (a), $\mathsf{rn}(\beta) = \beta \in \alpha$, hence

$$\beta = \mathsf{rn}(\beta) \in \bigcup \{ \mathsf{rn}(x) + 1 \mid \mathsf{rn}(x) \in \alpha \} = \mathsf{rn}(V_\alpha).$$

This completes the proof. □

We finally show that a class $\mathcal{A}$ is a set if and only if the ranks of their elements can be bounded by an ordinal.

LEMMA 3.39. *$\mathcal{A}$ is set iff there is an α such that $\forall y{\in}\mathcal{A}\,(\mathsf{rn}(y) \in \alpha)$.*

PROOF. $\rightarrow$. Let $\mathcal{A} = x$. From Lemma 3.37(a) we obtain that $\mathsf{rn}(x)$ is the α we need.

$\leftarrow$. Assume $\mathsf{rn}(x) \in \alpha$ for all $y \in \mathcal{A}$. Then $\mathcal{A} \subseteq \{ y \mid \mathsf{rn}(y) \in \alpha \} = V_\alpha$. □

4. Cardinals

We now introduce cardinals and develop their basic properties.

4.1. Size Comparison Between Sets. Define

$$|a| \leq |b| :\leftrightarrow \exists f.f\colon a \to b \text{ and } f \text{ injective},$$
$$|a| = |b| :\leftrightarrow \exists f.f\colon a \leftrightarrow b,$$
$$|a| < |b| :\leftrightarrow |a| \leq |b| \wedge |a| \neq |b|,$$
$$^b a := \{ f \mid f\colon b \to a \}.$$

Two sets a and b are *equinumerous* if $|a| = |b|$. Notice that we did not define $|a|$, but only the relations $|a| \leq |b|$, $|a| = |b|$ and $|a| < |b|$.

The following properties are clear:

$$|a \times b| = |b \times a|;$$
$$|^a(^b c)| = |^{a\times b} c|;$$
$$|\mathcal{P}(a)| = |^a\{0,1\}|.$$

THEOREM 4.1 (Cantor). $|a| < |\mathcal{P}(a)|$.

PROOF. Clearly $f\colon a \to \mathcal{P}(a)$, $x \mapsto \{x\}$ is injective. Assume that we have $g\colon a \leftrightarrow \mathcal{P}(a)$. Consider

$$b := \{ x \mid x \in a \wedge x \notin g(x) \}.$$

Then $b \subseteq a$, hence $b = g(x_0)$ for some $x_0 \in a$. It follows that $x_0 \in g(x_0) \leftrightarrow x_0 \notin g(x_0)$ and hence a contradiction. □

THEOREM 4.2 (Cantor, Bernstein). *If $a \subseteq b \subseteq c$ and $|a| = |c|$, then $|b| = |c|$.*

PROOF. Let $f\colon c \to a$ be bijective and $r := c \setminus b$. We recursively define $g\colon \omega \to V$ by

$$g(0) = r,$$
$$g(n+1) = f[g(n)].$$

Let

$$\overline{r} := \bigcup_n g(n)$$

and define $i\colon c \to b$ by

$$i(x) := \begin{cases} f(x), & \text{if } x \in \overline{r}, \\ x, & \text{if } x \notin \overline{r}. \end{cases}$$

It suffices to show that (a) $\mathsf{rng}(i) = b$ and (b) i is injective. Ad (a). Let $x \in b$. We must show $x \in \mathsf{rng}(i)$. Wlog let $x \in \overline{r}$. Because of $x \in b$ we then have $x \notin g(0)$. Hence there is an n such that $x \in g(n+1) = f[g(n)]$, so $x = f(y) = i(y)$ for some $y \in \overline{r}$. Ad (b). Let $x \neq y$. Wlog $x \in \overline{r}$, $y \notin \overline{r}$. But then $i(x) \in \overline{r}$, $i(y) \notin \overline{r}$, hence $i(x) \neq i(y)$. □

Remark. The Theorem of Cantor and Bernstein can be seen as an application of the Fixed Point Theorem of Knaster-Tarski.

Corollary 4.3. $|a| \leq |b| \to |b| \leq |a| \to |a| = |b|$.

Proof. Let $f\colon a \to b$ and $g\colon b \to a$ injective. Then $(g \circ f)[a] \subseteq g[b] \subseteq a$ and $|(g \circ f)[a]| = |a|$. By the Theorem of Cantor and Bernstein $|b| = |g[b]| = |a|$. □

4.2. Cardinals, Aleph Function. A cardinal is defined to be an ordinal that is not equinumerous to a smaller ordinal:

$$\alpha \text{ is a } \textit{cardinal} \text{ if } \forall \beta{<}\alpha\, (|\beta| \neq |\alpha|).$$

Here and later we write - because of Lemma 3.22 - $\alpha < \beta$ for $\alpha \in \beta$ and $\alpha \leq \beta$ for $\alpha \subseteq \beta$.

Lemma 4.4. $|n| = |m| \to n = m$.

Proof. Induction on n. 0. Clear. $n+1$. Let $f\colon n+1 \leftrightarrow m+1$. We may assume $f(n) = m$. Hence $f{\restriction}n\colon n \leftrightarrow m$ and therefore $n = m$ by IH, hence also $n+1 = m+1$. □

Corollary 4.5. *n is a cardinal.*

Lemma 4.6. $|n| \neq |\omega|$.

Proof. Assume $|n| = |\omega|$. Because of $n \subseteq n+1 \subseteq \omega$ the Theorem of Cantor and Bernstein implies $|n| = |n+1|$, a contradiction. □

Corollary 4.7. *ω is a cardinal.*

Lemma 4.8. $\omega \leq \alpha \to |\alpha + 1| = |\alpha|$.

Proof. Define $f\colon \alpha \to \alpha + 1$ by

$$f(x) := \begin{cases} \alpha, & \text{if } x = 0; \\ n, & \text{if } x = n+1; \\ x, & \text{otherwise.} \end{cases}$$

Then $f\colon \alpha \leftrightarrow \alpha + 1$. □

Corollary 4.9. *If $\omega \leq \alpha$ and α is a cardinal, then α is a limit.*

Proof. Assume $\alpha = \beta + 1$. Then $\omega \leq \beta < \alpha$, hence $|\beta| = |\beta + 1|$, contradicting the assumption that α is a cardinal. □

LEMMA 4.10. *If a is a set of cardinals, then* $\sup(a)$ *(*$:= \bigcup a$*) is a cardinal.*

PROOF. Otherwise there would be an $\alpha < \sup(a)$ such that $|\alpha| = |\sup(a)|$. Hence $\alpha \in \bigcup a$ and therefore $\alpha \in \beta \in a$ for some cardinal β. By the Theorem of Cantor and Bernstein from $\alpha \subseteq \beta \subseteq \bigcup a$ and $|\alpha| = |\bigcup a|$ it follows that $|\alpha| = |\beta|$. Because of $\alpha \in \beta$ and β a cardinal this is impossible. □

We now show that for every ordinal there is a strictly bigger cardinal. More generally, even the following holds:

THEOREM 4.11 (Hartogs).

$$\forall a \exists! \alpha . \forall \beta < \alpha (|\beta| \leq |a|) \wedge |\alpha| \not\leq |a|.$$

α is the Hartogs number *of a; it is denoted by $\mathcal{H}(a)$.*

PROOF. Uniqueness. Clear. Existence. Let $w := \{ (b, r) \mid b \subset a \wedge r \text{ well-ordering on } b \}$ and $\gamma_{(b,r)}$ the uniquely determined ordinal isomorphic to (b, r). Then $\{ \gamma_{(b,r)} \mid (b, r) \in w \}$ is a transitive subset of On, hence an ordinal α. We must show

(a) $\beta < \alpha \to |\beta| \leq |a|$,
(b) $|\alpha| \not\leq |a|$.

(a). Let $\beta < \alpha$. Then β is isomorphic to a $\gamma_{(b,r)}$ with $(b, r) \in w$, hence there exists an $f : \beta \leftrightarrow b$.

(b). Assume $f : \alpha \to a$ is injective. Then $\alpha = \gamma_{(b,r)}$ for some $b \subseteq a$ ($b := \mathsf{rng}(f)$), hence $\alpha \in \alpha$, a contradiction. □

REMARK. (a). The Hartogs number of a is a cardinal. For let α be the Hartogs number of a, $\beta < \alpha$. If $|\beta| = |\alpha|$, we would have $|\alpha| = |\beta| \leq |a|$, a contradiction.

(b). The Hartogs number of β is the least cardinal α such that $\alpha > \beta$.

The aleph function $\aleph : \mathsf{On} \to V$ is defined recursively by

$$\begin{aligned} \aleph_0 &:= \omega, \\ \aleph_{\alpha+1} &:= \mathcal{H}(\aleph_\alpha), \\ \aleph_\alpha &:= \sup\{ \aleph_\beta \mid \beta < \alpha \} \text{ for } \alpha \text{ limit.} \end{aligned}$$

LEMMA 4.12 (Properties of $\aleph$). (a) *$\aleph_\alpha$ is a cardinal.*
(b) $\alpha < \beta \to \aleph_\alpha < \aleph_\beta$.
(c) $\forall \beta . \beta$ *cardinal* $\to \omega \leq \beta \to \exists \alpha (\beta = \aleph_\alpha)$.

PROOF. (a). Induction on α; clear. (b). Induction on β. 0. Clear. $\beta + 1$.

$$\begin{aligned} &\alpha < \beta + 1 \\ &\alpha < \beta \vee \alpha = \beta \\ &\aleph_\alpha < \aleph_\beta \vee \aleph_\alpha = \aleph_\beta \\ &\aleph_\alpha < \aleph_{\beta+1}. \end{aligned}$$

β limit.

$$\begin{aligned} &\alpha < \beta \\ &\alpha < \gamma \text{ for some } \gamma < \beta \end{aligned}$$

$$\aleph_\alpha < \aleph_\gamma \leq \aleph_\beta.$$

(c). Let α be minimal such that $\beta \leq \aleph_\alpha$. Such an α exists, for otherwise $\aleph\colon \mathsf{On} \to \beta$ would be injective. We show $\aleph_\alpha \leq \beta$ by cases on α. 0. Clear. $\alpha = \alpha' + 1$. By the choice of α we have $\aleph_{\alpha'} < \beta$, hence $\aleph_\alpha \leq \beta$. α limit. By the choice of α we have $\aleph_\gamma < \beta$ for all $\gamma < \alpha$, hence $\aleph_\alpha = \sup\{\, \aleph_\gamma \mid \gamma < \alpha \,\} \leq \beta$. □

We show that every infinite ordinal is equinumerous to a cardinal.

LEMMA 4.13. $(\forall \beta \geq \omega)\exists \alpha(|\beta| = |\aleph_\alpha|)$.

PROOF. Consider $\delta := \min\{\, \gamma \mid \gamma < \beta \wedge |\gamma| = |\beta| \,\}$. Clearly δ is a cardinal. Moreover $\delta \geq \omega$, for otherwise

$$\begin{aligned} &\delta = n \\ &|n| = |\beta| \\ &n \subseteq n+1 \subseteq \beta \\ &|n| = |n+1|, \end{aligned}$$

a contradiction. Hence $\delta = \aleph_\alpha$ for some α, and therefore $|\delta| = |\beta| = |\aleph_\alpha|$. □

4.3. Products of Cardinals. We now show that $|\aleph_\alpha \times \aleph_\alpha| = |\aleph_\alpha|$. On the set $\mathsf{On} \times \mathsf{On}$ we define a relation $\prec$ by

$$\begin{aligned} (\alpha,\beta) \prec (\gamma,\delta) :\leftrightarrow\ & \max\{\alpha,\beta\} < \max\{\gamma,\delta\} \vee \\ & (\max\{\alpha,\beta\} = \max\{\gamma,\delta\} \wedge \alpha < \gamma) \vee \\ & (\max\{\alpha,\beta\} = \max\{\gamma,\delta\} \wedge \alpha = \gamma \wedge \beta < \delta). \end{aligned}$$

LEMMA 4.14. *$\prec$ is a well-ordering on $\mathsf{On} \times \mathsf{On}$.*

PROOF. Clearly $\prec$ is a linear ordering. To see the well-foundedness of $\prec$ consider an $a \subseteq \mathsf{On} \times \mathsf{On}$ such that $a \neq \emptyset$. Then

$$\emptyset \neq \mathcal{A} := \{\, \alpha \mid \exists \rho,\mu((\rho,\mu) \in a \wedge \max\{\rho,\mu\} = \alpha) \,\} \subseteq \mathsf{On}.$$

Let $\alpha_0 := \min(\mathcal{A})$. Then

$$\emptyset \neq \mathcal{A}_1 := \{\, \rho \mid \exists \mu((\rho,\mu) \in a \wedge \max\{\rho,\mu\} = \alpha_0) \,\} \subseteq \mathsf{On}.$$

Let $\rho_0 := \min(\mathcal{A}_1)$. Then

$$\emptyset \neq \mathcal{A}_2 := \{\, \mu \mid (\rho_0,\mu) \in a \wedge \max\{\rho_0,\mu\} = \alpha_0) \,\} \subseteq \mathsf{On}.$$

Let $\mu_0 := \min(\mathcal{A}_2)$. Then clearly $(\rho_0,\mu_0) = \min_\prec(a)$. Finally notice that $\widehat{(\alpha,\beta)}$ must be a set, for $\widehat{(\alpha,\beta)} \subseteq \gamma \times \gamma$ with $\gamma := \max\{\alpha,\beta\} + 1$. □

COROLLARY 4.15. *$\mathsf{On} \times \mathsf{On}$ is isomorphic to On (w.r.t. $\prec$ and $\in \restriction \mathsf{On}$).*

PROOF. By Lemma 4.14 $\prec$ is a well-ordering on $\mathsf{On} \times \mathsf{On}$. Hence by Corollary 3.19 there is an isomorphism onto a transitive and hence also ordinal class. This class cannot possibly be a set, for then $\mathsf{On} \times \mathsf{On}$ would be a set as well. But by Lemma 3.23(c) On is the only proper ordinal class. □

THEOREM 4.16. *$\aleph_\alpha \times \aleph_\alpha$ is isomorphic to $\aleph_\alpha$ (w.r.t. the relations $\prec$ on $\aleph_\alpha \times \aleph_\alpha$ and $\in$ on $\aleph_\alpha$).*

PROOF. Assume: $\exists\alpha(\aleph_\alpha \times \aleph_\alpha$ not isomorphic to $\aleph_\alpha)$. Let

$$\alpha_0 := \min\{\, \alpha \mid \aleph_\alpha \times \aleph_\alpha \text{ not isomorphic to } \aleph_\alpha \,\}.$$

Clearly $\alpha_0 \neq 0$. Since $\aleph_{\alpha_0} \times \aleph_{\alpha_0}$ and $\aleph_{\alpha_0}$ are well-ordered sets, one of them must be isomorphic to a proper initial segment of the other. Therefore we distinguish two cases.

Case (a). $\aleph_{\alpha_0}$ is isomorphic to $\widehat{(\beta,\gamma)}$ with $\beta,\gamma < \aleph_{\alpha_0}$. Choose $\delta < \aleph_{\alpha_0}$ with $\beta,\gamma < \delta$. Then $\widehat{(\beta,\gamma)} \subseteq \delta \times \delta$, and

$$\begin{aligned} |\aleph_{\alpha_0}| = |\widehat{(\beta,\gamma)}| \leq |\delta \times \delta| = |\aleph_\tau \times \aleph_\tau| \quad & \text{for some } \tau < \alpha_0 \\ = |\aleph_\tau| \quad & \text{by choice of } \alpha_0, \end{aligned}$$

hence a contradiction to Lemma 4.12(b).

Case (b). $\aleph_{\alpha_0} \times \aleph_{\alpha_0}$ is isomorphic to $\beta < \aleph_{\alpha_0}$. Then

$$\begin{aligned} &|\aleph_{\alpha_0}| \leq |\aleph_{\alpha_0} \times \aleph_{\alpha_0}| = |\beta| \leq |\aleph_{\alpha_0}| \\ &|\aleph_{\alpha_0}| = |\beta|, \end{aligned}$$

hence a contradiction to the fact that $\aleph_{\alpha_0}$ is a cardinal, $\beta < \aleph_{\alpha_0}$. □

COROLLARY 4.17. (a) $|\aleph_\alpha \times \aleph_\beta| = |\max\{\aleph_\alpha, \aleph_\beta\}|$.
(b) $n \neq 0 \to |{}^n\aleph_\alpha| = |\aleph_\alpha|$.

PROOF. (a). We may assume $\alpha \leq \beta$. Then

$$|\aleph_\beta| \leq |\aleph_\alpha \times \aleph_\beta| \leq |\aleph_\beta \times \aleph_\beta| = |\aleph_\beta|.$$

(b). This follows easily from Theorem 4.16, by induction on n. □

5. The Axiom of Choice

5.1. Axiom of Choice, Well Ordering Theorem, Zorn's Lemma. A relation $\mathcal{R}$ on $\mathcal{A}$ is a *partial ordering* if for all $x, y, z \in \mathcal{A}$

$$\begin{aligned} &\neg x\mathcal{R}x, && \text{irreflexivity} \\ &x\mathcal{R}y \to y\mathcal{R}z \to x\mathcal{R}z, && \text{transitivity.} \end{aligned}$$

An element $x \in \mathcal{A}$ is *maximal* if there is no $y \in \mathcal{A}$ such that $x\mathcal{R}y$. Let $\mathcal{B} \subseteq \mathcal{A}$. An element $x \in \mathcal{A}$ is an *upper bound* of $\mathcal{B}$ if

$$\forall y{\in}\mathcal{B}.y\mathcal{R}x \vee y = x.$$

THEOREM 5.1. *The following are equivalent.*

(a) *The* axiom of choice (AC)

$$\forall x.\emptyset \notin x \to \exists f.f\colon x \to \bigcup x \wedge (\forall y{\in}x)(f(y) \in y).$$

(b) *The* well ordering theorem (WO)

$$\forall a \exists r(r \textit{ is a well ordering on } a).$$

(c) Zorn's Lemma (ZL)*: Let $(P,<)$ be a non empty partial ordering, with the property that every (by $<$) linearly ordered subset $L \subseteq P$ has an upper bound in P. Then P has a maximal element.*

PROOF. (ZL) $\to$ (WO). Let a be given, and define

$$P := \{\, f \mid \exists\alpha(f\colon \alpha \to a \text{ injective}) \,\} \subseteq \mathcal{P}(\mathcal{H}(a) \times a).$$

P is partially ordered by proper inclusion $\subsetneq$. Let $L \subseteq P$ be linearly ordered. Then $\bigcup L \in P$, hence $\bigcup L$ is an upper bound of L. Zorn's Lemma then gives a maximal element $f_0 \in P$. Clearly f_0 is a bijection of an ordinal α_0 onto a, hence f_0 induces a well ordering on a.

(WO) $\to$ (AC). Let $\emptyset \notin x$. By (WO) there is a well ordering $<$ on $\bigcup x$. Clearly $<$ induces a well ordering on every $y \in x$. Define

$$\begin{aligned} f\colon x &\to \bigcup x, \\ y &\mapsto \min{}_<(y) \in y. \end{aligned}$$

(AC) $\to$ (ZL). Let $<$ be a partial ordering on $P \neq \emptyset$. Assume that every subset $L \subseteq P$ linearly ordered by $<$ has an upper bound in P. By (AC) there is a choice function f on $\mathcal{P}(P) \setminus \{\emptyset\}$. Let $z \notin P$ be arbitrary, and define

$$\mathcal{F}\colon \mathsf{On} \to V$$

$$\mathcal{F}(\alpha) = \begin{cases} f(\{\, y \mid y \in P \setminus \mathcal{F}[\alpha] \wedge y \text{ upper bound of } \mathcal{F}[\alpha] \,\}), & \text{if } \{\dots\} \neq \emptyset; \\ z, & \text{otherwise.} \end{cases}$$

Then there is a ρ such that $\mathcal{F}(\rho) = z$, for otherwise $\mathcal{F}\colon \mathsf{On} \to P$ would be injective, contradicting our assumption that P is a set. Let $\rho_0 := \min\{\, \rho \mid \mathcal{F}(\rho) = z \,\}$. $\mathcal{F}[\rho_0]$ is linearly ordered, and we have $\mathcal{F}[\rho_0] \subseteq P$. By assumption there is an upper bound $y_0 \in P$ of $\mathcal{F}[\rho_0]$. We show that y_0 is a maximal element in P. So assume $y_0 < y$ for some $y \in P$. Then y is an upper bound of $\mathcal{F}[\rho_0]$ and $y \notin \mathcal{F}[\rho_0]$. But this contradicts the definition of ρ_0. □

From now on we will assume the axiom of choice; however, we will mark every theorem and every definition depending on it by (AC).

(AC) clearly is equivalent to its special case where every two elements $y_1, y_2 \in x$ are disjoint. We hence note the following equivalent to the axiom of choice:

LEMMA 5.2. *The following are equivalent*

(a) *The axiom of choice (AC).*
(b) *For every surjective $g\colon a \to b$ there is an injective $f\colon b \to a$ such that $(\forall x{\in}b)(g(fx) = x)$.*

PROOF. (a) $\Rightarrow$ (b). Let $g\colon b \to a$ surjective. By (AC) there is a well-ordering $<$ of b. Define $f\colon a \to b$ by $f(x) := \min_<\{\, y \mid y \in b \wedge g(y) = x \,\}$.

(b) $\Rightarrow$ (a). We may assume $x \neq \emptyset$ and $(\forall y_1, y_2 \in x)(y_1 \cap y_2 = \emptyset)$. Define $g\colon \bigcup x \to x$ by $g(z) :=$ the unique $y \in x$ such that $z \in y$. Then g is surjective. By (b) there is an injective $f\colon x \to \bigcup x$ such that $g(fy) = y$ for all $y \in x$, hence $f(y) \in y$. □

5.2. Cardinality. α is the *cardinality* of a if α is a cardinal and there is a bijection $f\colon a \to \alpha$.

THEOREM 5.3 (AC). *Every set has a unique cardinality.*

PROOF. Uniqueness. Clear. Existence. Let $<$ be a well–ordering on a. Then there is a γ such that a is isomorphic to γ. Hence $\{\,\tau \mid |\tau| = |a|\,\} \neq \emptyset$ and therefore $\min\{\,\tau \mid |\tau| = |a|\,\}$ is a cardinal. □

Clearly $|a| = |b|$ iff the cardinality of a equals the cardinality of b, and $|a| \leq |b|$ iff the cardinality of a is less than or equal to the cardinality of b. Therefore we can use $|a|$ as a notation for the cardinality of a.

A set a is defined to be *finite* if a can be mapped bijectively onto a natural number, and *infinite* otherwise. Using (AC) it follows that a is finite iff $|a| < \omega$.

LEMMA 5.4 (AC). *If $a, b \neq \emptyset$ and a or b is infinite, then*

$$|a \times b| = \max\{|a|, |b|\}.$$

PROOF. Let $|a| = \max\{|a|, |b|\}$. Then

$$|a| \leq |a \times b| = ||a| \times |b|| \leq ||a| \times |a|| = |a|.$$

□

THEOREM 5.5 (AC). *Let I be infinite or $\sup_{i\in I} |A_i|$ be infinite. Then*

(a) $|\bigcup_{i\in I} A_i| \leq \max\{|I|, \sup_{i\in I} |A_i|\}$.
(b) *If in addition $(\forall i \in I)(A_i \neq \emptyset)$ and $(\forall i, j \in I)(i \neq j \to A_i \cap A_j = \emptyset)$, then equality holds.*

PROOF. (a). We may assume $\kappa := \sup_{i\in I} |A_i| \neq 0$. Choose a well-ordering $<$ of I and define w.r.t. this well-ordering

$$f\colon \bigcup_{i\in I} A_i \to \bigcup_{i\in I} (\{i\} \times A_i),$$
$$f(x) = (\min\{\, i \in I \mid x \in A_i \,\}, x).$$

Clearly f is injective. Hence

$$\begin{aligned} |\bigcup_{i\in I} A_i| &\leq |\bigcup_{i\in I} (\{i\} \times A_i)| \\ &\leq |\bigcup_{i\in I} (\{i\} \times \kappa)| \\ &= |I \times \kappa| \\ &= \max\{|I|, \kappa\}. \end{aligned}$$

(b). Because of (a) it suffices to show that $|I|, |A_i| \leq |\bigcup_{i\in I} A_i|$. The second estimate is clear. For the first one choose a well-ordering $<$ of $\bigcup_{i\in I} A_i$ and define $f\colon I \to \bigcup_{i\in I} A_i$ by $f(i) := \min_<\{\, x \mid x \in A_i \,\}$. By our assumption f is injective. □

A set a is *Dedekind-finite* if a cannot be mapped bijectively onto a proper subset b of a, otherwise *Dedekind-infinite*.

THEOREM 5.6 (AC). *A set a is Dedekind-infinite iff a is infinite.*

PROOF. $\to$. Let $b \subsetneq a$ and $f\colon a \leftrightarrow b$. Assume $|a| < \omega$, say $|a| = n$. Then there is a $c \subsetneq n$ and some $g\colon n \leftrightarrow c$. We show by induction on n that this is impossible:

$$\forall n \neg(\exists c \subsetneq n) \exists g(g\colon n \leftrightarrow c).$$

0. Clear. $n{+}1$. Let $g\colon n{+}1 \leftrightarrow c$ and $c \subsetneq n{+}1$. We may assume $n \notin \mathsf{rng}(g{\restriction}n)$. It follows that $g{\restriction}n\colon n \leftrightarrow c \setminus \{n\} \subsetneq n$ and hence a contradiction to the IH.

$\leftarrow$. Let $g\colon \omega \to a$ be injective and $h\colon g[\omega] \leftrightarrow g[\omega \setminus 1]$ defined by

$$h = \{\, (g(n), g(n+1)) \mid n \in \omega \,\}.$$

Define $f\colon a \leftrightarrow (a \setminus \{g(0)\})$ by

$$f(x) = \begin{cases} x, & \text{if } x \in a \setminus g[\omega]; \\ h(x), & \text{otherwise.} \end{cases}$$

□

5.3. Regular and Singular Cardinals. Let κ, λ denote cardinals $\geq \omega$. In this section we shall always assume the Axiom of Choice (AC).

Definition 5.7 (AC). (a) $x \subseteq \kappa$ is *confinal* in κ if $\sup(x) = \kappa$.
(b) $\mathsf{cf}(\kappa) := \min\{\, |x| \mid x \subseteq \kappa$ and x confinal with $\kappa \,\}$ is the *confinality* of κ.
(c) κ is *regular* if $\mathsf{cf}(\kappa) = \kappa$.
(d) κ is *singular* if $\mathsf{cf}(\kappa) < \kappa$.

Theorem 5.8 (AC). (a) $\omega = \aleph_0$ *is regular.*
(b) $\aleph_{\alpha+1}$ *is regular.*
(c) *If β is a limit and $\beta < \aleph_\beta$, then $\aleph_\beta$ is singular.*

Proof. (a). Assume ω is singular, that is $\mathsf{cf}(\omega) < \omega$. Then there is an $x \subseteq \omega$ such that $|x| = n$ and $\sup(x) = \omega$. But this is impossible (proof by induction on n).

(b). Assume $\aleph_{\alpha+1}$ is singular. Then $\mathsf{cf}(\aleph_{a+1}) \leq \aleph_a$. Hence there is an $x \subseteq \aleph_{\alpha+1}$ such that $|x| \leq \aleph_\alpha$ and $\sup(x) = \aleph_{\alpha+1}$. But then

$$\begin{aligned} \aleph_{\alpha+1} &= |\bigcup x| \\ &\leq \max\{|x|, \sup\{\, |y| \mid y \in x \,\}\} \quad \text{by Theorem 5.5(a)} \\ &\leq \aleph_\alpha, \end{aligned}$$

a contradiction.

(c). Let β be a limit such that $\beta < \aleph_\beta$. Then we have $\aleph_\beta = \sup\{\, \aleph_\gamma \mid \gamma < \beta \,\}$ and moreover $|\{\, \aleph_\gamma \mid \gamma < \beta \,\}| = |\beta| < \aleph_\beta$. Hence $\aleph_\beta$ is singular. □

By definition for every infinite cardinal κ there is a subset $x \subseteq \kappa$ whose cardinality equals $\mathsf{cf}(\kappa)$, hence which can be mapped bijectively onto $\mathsf{cf}(\kappa)$. We now show that one can even assume that this bijection is an isomorphism.

Lemma 5.9 (AC). *Let κ be an infinite cardinal. Then there exists a subset $x \subseteq \kappa$ confinal in κ that is isomorphic to $\mathsf{cf}(\kappa)$.*

Proof. Let $y \subseteq \kappa$, $\sup(y) = \kappa$, $|y| = \mathsf{cf}(\kappa)$ and $g\colon \mathsf{cf}(\kappa) \leftrightarrow y$. By transfinite recursion we define

$$\begin{aligned} &\mathcal{F}\colon \mathsf{On} \to V, \\ &\mathcal{F}(\alpha) := \sup(\mathcal{F}[\alpha] \cup g[\alpha]) + 1. \end{aligned}$$

Let $f := \mathcal{F}{\restriction}\mathsf{cf}(\kappa)$. One can see easily
(a) $\alpha < \beta < \mathsf{cf}(\kappa) \to f(\alpha) < f(\beta) \wedge g(\alpha) < f(\beta)$.
(b) $\mathsf{rng}(f) \subseteq \kappa$
(c) $\mathsf{rng}(f)$ is confinal with κ.

$\mathsf{rng}(f)$ is the x we are looking for. □

COROLLARY 5.10 (AC). *If κ is an infinite cardinal, then $\mathsf{cf}(\kappa)$ is a regular cardinal.*

PROOF. $\mathsf{cf}(\mathsf{cf}(\kappa)) \leq \mathsf{cf}(\kappa)$ is clear. We must show $\mathsf{cf}(\kappa) \leq \mathsf{cf}(\mathsf{cf}(\kappa))$. By the lemma above there are x, f such that $x \subseteq \kappa$, $\sup(x) = \kappa$ and $f\colon \mathsf{cf}(\kappa) \leftrightarrow x$ isomorphism. Moreover there is $y \subseteq \mathsf{cf}(\kappa)$ such that $\sup(y) = \mathsf{cf}(\kappa)$ and $|y| = \mathsf{cf}(\mathsf{cf}(\kappa))$. One can see easily that $\{\, f(\alpha) \mid \alpha \in y \,\}$ is confinal with κ. Hence

$$\begin{aligned}\mathsf{cf}(\kappa) &\leq |\{\, f(\alpha) \mid \alpha \in y \,\}| \\ &= |y| \\ &= \mathsf{cf}(\mathsf{cf}(\kappa))\end{aligned}$$

This concludes the proof. □

THEOREM 5.11 (König). *Let κ be an infinite cardinal. Then $\kappa < |^{\mathsf{cf}(\kappa)}\kappa|$.*

PROOF. $\kappa = |^1\kappa| \leq |^{\mathsf{cf}(\kappa)}\kappa|$ is clear. Hence it suffices to derive a contradiction from the assumption that there is a bijection $f\colon \kappa \leftrightarrow {}^{\mathsf{cf}(\kappa)}\kappa$. According to Lemma 5.9 there exists $x \subseteq \kappa$ such that $\sup(x) = \kappa$ and moreover an isomorphism $g\colon \mathsf{cf}(\kappa) \leftrightarrow x$. For every $\alpha < \mathsf{cf}(\kappa)$ we therefore have $g(\alpha) < \kappa$ and hence

$$|\{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\}| \leq |g(\alpha)| < \kappa,$$

hence $\{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\} \subsetneq \kappa$. Let

$$\begin{aligned}&h\colon \mathsf{cf}(\kappa) \to \kappa, \\ &h(\alpha) := \min(\kappa \setminus \{\, f(\gamma)(\alpha) \mid \gamma < g(\alpha) \,\}).\end{aligned}$$

We obtain the desired contradiction by showing that $f(\gamma) \neq h$ for all $\gamma < \kappa$. So let $\gamma < \kappa$. Choose $\alpha < \mathsf{cf}(\kappa)$ such that $\gamma < g(\alpha)$. Then $h(\alpha) \neq f(\gamma)(\alpha)$ by construction of h. □

5.4. Cardinal Powers, Continuum Hypothesis. In this section we again assume (AC). We define

$$\aleph_\alpha^{\aleph_\beta} := |^{\aleph_\beta}\aleph_\alpha|.$$

Later we will introduce powers of ordinals as well. It should always be clear from the context whether we mean ordinal or cardinal power.

THEOREM 5.12 (AC). (a) $\aleph_\beta < \mathsf{cf}(\aleph_\alpha) \to \aleph_\alpha \leq \aleph_\alpha^{\aleph_\beta} \leq |\mathcal{P}(\aleph_\alpha)|$
(b) $\mathsf{cf}(\aleph_\alpha) \leq \aleph_\beta \leq \aleph_\alpha \to \aleph_\alpha < \aleph_\alpha^{\aleph_\beta} \leq |\mathcal{P}(\aleph_\alpha)|$
(c) $\aleph_\alpha \leq \aleph_\beta \to \aleph_\alpha^{\aleph_\beta} = |\mathcal{P}(\aleph_\beta)|$.

PROOF. (a).

$$\begin{aligned}\aleph_\alpha &\leq |^{\aleph_\beta}\aleph_\alpha| \\ &\leq |^{\aleph_\beta}(^{\aleph_\alpha}\{0,1\})| \\ &= |^{\aleph_\beta \times \aleph_\alpha}\{0,1\}| \\ &= |^{\aleph_\alpha}\{0,1\}| \quad \text{because } \aleph_\beta \leq \aleph_\alpha \\ &= |\mathcal{P}(\aleph_\alpha)|.\end{aligned}$$

(b).

$$\begin{aligned}\aleph_\alpha &< |{}^{\mathsf{cf}(\aleph_\alpha)}\aleph_\alpha| && \text{König's Theorem}\\ &\leq |{}^{\aleph_\beta}\aleph_\alpha| \\ &\leq |\mathcal{P}(\aleph_\alpha)| && \text{as in (a).}\end{aligned}$$

(c).

$$\begin{aligned}|\mathcal{P}(\aleph_\beta)| &= |{}^{\aleph_\beta}\{0,1\}| \\ &\leq |{}^{\aleph_\beta}\aleph_\alpha| \\ &\leq |{}^{\aleph_\beta\times\aleph_\alpha}\{0,1\}| \\ &= |{}^{\aleph_\beta}\{0,1\}| \\ &= |\mathcal{P}(\aleph_\beta)|.\end{aligned}$$

This concludes the proof. □

One can say much more about cardinal powers if one assumes the so-called *continuum hypothesis*:

$$|\mathcal{P}(\aleph_0)| = \aleph_1. \tag{CH}$$

An obvious generalization to all cardinals is the *generalized continuum hypothesis*:

$$|\mathcal{P}(\aleph_\alpha)| = \aleph_{\alpha+1}. \tag{GCH}$$

It is an open problem whether the continuum hypothesis holds in the cumulative type structure (No. 1 in Hilbert's list of mathematical problems, posed in a lecture at the international congress of mathematicians in Paris 1900). However, it is known that continuum hypothesis is independent from the other axioms of set theory. We shall always indicate use of (CH) or (GCH).

THEOREM 5.13 (GCH). (a) $\aleph_\beta < \mathsf{cf}(\aleph_\alpha) \to \aleph_\alpha = \aleph_\alpha^{\aleph_\beta}$.
(b) $\mathsf{cf}(\aleph_\alpha) \leq \aleph_\beta \leq \aleph_\alpha \to \aleph_\alpha^{\aleph_\beta} = \aleph_{\alpha+1}$.
(c) $\aleph_\alpha \leq \aleph_\beta \to \aleph_\alpha^{\aleph_\beta} = \aleph_{\beta+1}$.

PROOF. (b) and (c) follow with (GCH) from the previous theorem.
(a). Let $\aleph_\beta < \mathsf{cf}(\aleph_\alpha)$. First note that

$${}^{\aleph_\beta}\aleph_\alpha = \bigcup\{\, {}^{\aleph_\beta}\gamma \mid \gamma < \aleph_\alpha \,\}$$

This can be seen as follows. $\supseteq$ is clear. $\subseteq$. Let $f\colon \aleph_\beta \to \aleph_\alpha$. Because of $|f[\aleph_\beta]| \leq \aleph_\beta < \mathsf{cf}(\aleph_\alpha)$ we have $\sup(f[\aleph_\beta]) < \gamma < \aleph_\alpha$ for some γ, hence $f\colon \aleph_\beta \to \gamma$.

This gives

$$\begin{aligned}\aleph_\alpha &\leq |{}^{\aleph_\beta}\aleph_\alpha| && \text{previous theorem}\\ &= |\bigcup\{\, {}^{\aleph_\beta}\gamma \mid \gamma < \aleph_\alpha \,\}| && \text{by the note above}\\ &\leq \max\{|\aleph_\alpha|, \sup_{\gamma<\aleph_\alpha} |{}^{\aleph_\beta}\gamma|\,\} && \text{by Theorem 5.5(a)}\end{aligned}$$

Hence it suffices to show that $|{}^{\aleph_\beta}\gamma| \leq \aleph_\alpha$ for $\gamma < \aleph_\alpha$. So let $\gamma < \aleph_\alpha$.

$$|{}^{\aleph_\beta}\gamma| \leq |{}^{\aleph_\beta\times\gamma}\{0,1\}|$$

$$\begin{aligned}
&\leq |^{\aleph_\beta \times \aleph_\delta}\{0,1\}| \quad \text{for some } \delta \text{ with } |\gamma| \leq \aleph_\delta < \aleph_\alpha \\
&\leq \begin{cases} |\mathcal{P}(\aleph_\delta)| & \text{if } \beta < \delta \\ |\mathcal{P}(\aleph_\beta)| & \text{if } \delta \leq \beta \end{cases} \\
&= \begin{cases} \aleph_{\delta+1} & \text{if } \beta < \delta \\ \aleph_{\beta+1} & \text{if } \delta \leq \beta \end{cases} \\
&\leq \aleph_\alpha.
\end{aligned}$$

This concludes the proof. □

6. Ordinal Arithmetic

We define addition, multiplication and exponentiation for ordinals and prove their basic properties. We also treat Cantor's normal form.

6.1. Ordinal Addition. Let

$$\begin{aligned}
&\alpha + 0 := \alpha, \\
&\alpha + (\beta + 1) := (\alpha + \beta) + 1, \\
&\alpha + \beta := \sup\{\, \alpha + \gamma \mid \gamma < \beta \,\} \quad \text{if } \beta \text{ limit.}
\end{aligned}$$

More precisely, define $s_\alpha \colon \mathsf{On} \to V$ by

$$\begin{aligned}
&s_\alpha(0) := \alpha, \\
&s_\alpha(\beta + 1) := s_\alpha(\beta) + 1, \\
&s_\alpha(\beta) := \bigcup \mathsf{rng}(s_\alpha \restriction \beta) \quad \text{if } \beta \text{ limit}
\end{aligned}$$

and then let $\alpha + \beta := s_\alpha(\beta)$.

Lemma 6.1 (Properties of Ordinal Addition). (a) $\alpha + \beta \in \mathsf{On}$.
(b) $0 + \beta = \beta$.
(c) $\exists \alpha, \beta (\alpha + \beta \neq \beta + \alpha)$.
(d) $\beta < \gamma \to \alpha + \beta < \alpha + \gamma$.
(e) *There are* α, β, γ *such that* $\alpha < \beta$, *but* $\alpha + \gamma \not< \beta + \gamma$.
(f) $\alpha \leq \beta \to \alpha + \gamma \leq \beta + \gamma$.
(g) *For* $\alpha \leq \beta$ *there is a unique* γ *such that* $\alpha + \gamma = \beta$.
(h) *If* β *is a limit, then so is* $\alpha + \beta$.
(i) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

Proof. (a). Induction on β. **Case** 0. Clear. **Case** $\beta + 1$. Then $\alpha + (\beta + 1) = (\alpha + \beta) + 1 \in \mathsf{On}$, for by IH $\alpha + \beta \in \mathsf{On}$. **Case** β limit. Then $\alpha + \beta = \sup\{\, \alpha + \gamma \mid \gamma < \beta \,\} \in \mathsf{On}$, for by IH $\alpha + \gamma \in \mathsf{On}$ for all $\gamma < \beta$.

(b). Induction on β. **Case** 0. Clear. **Case** $\beta + 1$. Then $0 + (\beta + 1) = (0 + \beta) + 1 = \beta + 1$, for by IH $0 + \beta = \beta$. **Case** β limit. Then

$$\begin{aligned}
0 + \beta &= \sup\{\, 0 + \gamma \mid \gamma < \beta \,\} \\
&= \sup\{\, \gamma \mid \gamma < \beta \,\} && \text{by IH} \\
&= \bigcup \beta \\
&= \beta, && \text{because } \beta \text{ is a limit.}
\end{aligned}$$

(c). $1 + \omega = \sup\{\, 1 + n \mid n \in \omega \,\} = \omega \neq \omega + 1$.

(d). Induction on γ. **Case** 0. Clear. **Case** $\gamma+1$. Then

$$\begin{aligned}
&\beta < \gamma+1,\\
&\beta < \gamma \vee \beta = \gamma,\\
&\alpha+\beta < \alpha+\gamma \vee \alpha+\beta = \alpha+\gamma && \text{by IH},\\
&\alpha+\beta \leq \alpha+\gamma < (\alpha+\gamma)+1 = \alpha+(\gamma+1).
\end{aligned}$$

Case γ limit. Let $\beta<\gamma$, hence $\beta<\delta$ for some $\delta<\gamma$. Then $\alpha+\beta<\alpha+\delta$ by IH, hence $\alpha+\beta<\sup\{\,\alpha+\delta \mid \delta<\gamma\,\}=\alpha+\gamma$.

(e). $0<1$, but $0+\omega=\omega=1+\omega$

(f). We first remark that there can be no β such that $\alpha<\beta<\alpha+1$, for otherwise we would have in case $\beta\in\alpha$ the contradiction $\beta\in\alpha\in\beta$ and in case $\beta=\alpha$ the contradiction $\alpha\in\alpha$. As a second preliminary remark we note that

$$\alpha\leq\beta\rightarrow\alpha+1\leq\beta+1,$$

for in case $\beta+1<\alpha+1$ we would have $\alpha<\beta+1<\alpha+1$, which cannot be the case (as we have just seen). – We now show the claim $\alpha\leq\beta\rightarrow\alpha+\gamma\leq\beta+\gamma$ by induction on γ. **Case** 0. Clear. **Case** $\gamma+1$. Then

$$\begin{aligned}
&\alpha+\gamma\leq\beta+\gamma && \text{by IH},\\
&(\alpha+\gamma)+1\leq(\beta+\gamma)+1 && \text{by the second preliminary remark},\\
&\alpha+(\gamma+1)\leq\beta+(\gamma+1) && \text{by definition}.
\end{aligned}$$

Case γ limit. Then

$$\begin{aligned}
&\alpha+\delta\leq\beta+\delta && \text{for all } \delta<\gamma, \text{ by IH},\\
&\alpha+\delta\leq\sup\{\,\beta+\delta \mid \delta<\gamma\,\}\\
&\sup\{\,\alpha+\delta \mid \delta<\gamma\,\}\leq\sup\{\,\beta+\delta \mid \delta<\gamma\,\}\\
&\alpha+\gamma\leq\beta+\gamma && \text{by definition}.
\end{aligned}$$

(g). Uniqueness of γ follows from (d). Existence: Let $\alpha\leq\beta$. By (b) and (f) $\beta=0+\beta\leq\alpha+\beta$. Let γ be the least ordinal such that $\beta\leq\alpha+\gamma$. We show that $\beta=\alpha+\gamma$. **Case** $\gamma=0$. Then $\beta\leq\alpha+\gamma=\alpha+0=\alpha\leq\beta$, hence $\beta=\alpha+\gamma$. **Case** $\gamma=\gamma'+1$. Then $\alpha+\gamma'<\beta$, hence $(\alpha+\gamma')+1\leq\beta$ by the first preliminary remark for (f) and hence $\alpha+\gamma=\beta$. **Case** γ limit. Then $\alpha+\delta<\beta$ for all $\delta<\gamma$, hence $\alpha+\gamma=\sup\{\,\alpha+\delta \mid \delta<\gamma\,\}\leq\beta$ and hence $\alpha+\gamma=\beta$.

(h). Let β limit. We use the characterization of limits in Lemma 3.27(a). $\alpha+\beta\neq 0$: Because of $0\leq\alpha$ we have $0<\beta=0+\beta\leq\alpha+\beta$ by (f). $\gamma<\alpha+\beta\rightarrow\gamma+1<\alpha+\beta$: Let $\gamma<\alpha+\beta=\sup\{\,\alpha+\delta \mid \delta<\beta\,\}$, hence $\gamma<\alpha+\delta$ for some $\delta<\beta$, hence $\gamma+1<\alpha+(\delta+1)$ with $\delta+1<\beta$, hence $\gamma+1<\sup\{\,\alpha+\delta \mid \delta<\beta\,\}$.

(i). Induction on γ. **Case** 0. Clear. **Case** $\gamma+1$. Then

$$\begin{aligned}
(\alpha+\beta)+(\gamma+1) &= [(\alpha+\beta)+\gamma]+1\\
&= [\alpha+(\beta+\gamma)]+1 \quad \text{by IH}\\
&= \alpha+[(\beta+\gamma)+1]\\
&= \alpha+[\beta+(\gamma+1)]
\end{aligned}$$

Case γ limit. By (h) also $\beta+\gamma$ is a limit. Hence

$$\begin{aligned}(\alpha+\beta)+\gamma &= \sup\{\,(\alpha+\beta)+\delta \mid \delta<\gamma\,\}\\ &= \sup\{\,\alpha+(\beta+\delta) \mid \delta<\gamma\,\} \quad \text{by IH}\\ &= \sup\{\,\alpha+\varepsilon \mid \varepsilon<\beta+\gamma\,\} \quad \text{see below}\\ &= \alpha+(\beta+\gamma).\end{aligned}$$

The equality of both suprema can be seen as follows. If $\varepsilon<\beta+\gamma$, then $\varepsilon<\beta+\delta$ for some $\delta<\gamma$ (by definition of $\beta+\gamma$) and hence $\alpha+\varepsilon<\alpha+(\beta+\delta)$. If conversely $\delta<\gamma$, then $\beta+\delta<\beta+\gamma$, hence $\alpha+(\beta+\delta)=\alpha+\varepsilon$ for some $\varepsilon<\beta+\gamma$ (take $\varepsilon:=\beta+\delta$). □

6.2. Ordinal Multiplication. Ordinal multiplication is defined by

$$\begin{aligned}&\alpha\cdot 0 := 0,\\ &\alpha\cdot(\beta+1) := (\alpha\cdot\beta)+\alpha,\\ &\alpha\cdot\beta := \sup\{\,\alpha\cdot\gamma \mid \gamma<\beta\,\} \quad \text{if } \beta \text{ limit.}\end{aligned}$$

We write $\alpha\beta$ for $\alpha\cdot\beta$.

LEMMA 6.2 (Properties of Ordinal Multiplication). (a) $\alpha\beta\in\mathsf{On}$.
(b) $0\beta=0$, $1\beta=\beta$.
(c) $\exists\alpha,\beta(\alpha\beta\neq\beta\alpha)$.
(d) $0<\alpha\to\beta<\gamma\to\alpha\beta<\alpha\gamma$.
(e) *There are* α,β,γ *such that* $0<\gamma$ *and* $\alpha<\beta$, *but* $\alpha\gamma\not<\beta\gamma$.
(f) $\alpha\leq\beta\to\alpha\gamma\leq\beta\gamma$.
(g) *If* $0<\alpha$ *and* β *is a limit, then so is* $\alpha\beta$.
(h) $\alpha(\beta+\gamma)=\alpha\beta+\alpha\gamma$.
(i) *There are* α,β,γ *such that* $(\alpha+\beta)\gamma\neq\alpha\gamma+\beta\gamma$.
(j) $\alpha\beta=0\to\alpha=0\vee\beta=0$.
(k) $(\alpha\beta)\gamma=\alpha(\beta\gamma)$.
(l) *If* $0<\beta$, *then there are unique* γ,ρ *such that* $\alpha=\beta\gamma+\rho$ *and* $\rho<\beta$.

PROOF. (a). Induction on β. **Case** 0. Clear. **Case** $\beta+1$. Then $\alpha(\beta+1)=(\alpha\beta)+\alpha\in\mathsf{On}$, for by IH $\alpha\beta\in\mathsf{On}$. **Case** β limit. Then $\alpha\beta=\sup\{\,\alpha\gamma \mid \gamma<\beta\,\}\in\mathsf{On}$, for by IH $\alpha\gamma\in\mathsf{On}$ for all $\gamma<\beta$.

(b). $0\beta=0$: Induction on β. **Case** 0. Clear. **Case** $\beta+1$. Then $0(\beta+1)=(0\beta)+0=0$ by IH. **Case** β limit. $0\beta=\sup\{\,0\gamma \mid \gamma<\beta\,\}=0$ by IH. – $1\beta=\beta$: Induction on β. **Case** 0. Clear. **Case** $\beta+1$. Then $1(\beta+1)=(1\beta)+1=\beta+1$ by IH. **Case** β limit. $1\beta=\sup\{\,1\gamma \mid \gamma<\beta\,\}=\sup\{\,\gamma \mid \gamma<\beta\,\}=\beta$ by IH.

(c). Fist note that for all $n\in\omega$ we have $n\omega=\sup\{\,nm \mid m<\omega\,\}=\omega$. This implies $2\omega=\omega$, but $\omega 2=\omega(1+1)=\omega 1+\omega=\omega+\omega>\omega$.

(d). Let $0<\alpha$. We show $\beta<\gamma\to\alpha\beta<\alpha\gamma$ by induction on γ. **Case** 0. Clear. **Case** $\gamma+1$. Then

$$\begin{aligned}&\beta<\gamma+1,\\ &\beta<\gamma\vee\beta=\gamma,\\ &\alpha\beta<\alpha\gamma\vee\alpha\beta=\alpha\gamma \quad \text{by IH,}\\ &\alpha\beta\leq\alpha\gamma<(\alpha\gamma)+\alpha=\alpha(\gamma+1).\end{aligned}$$

Case γ limit. Let $\beta < \gamma$, hence $\beta < \delta$ for some $\delta < \gamma$. Then $\alpha\beta < \alpha\delta$ by IH, hence $\alpha\beta < \sup\{\, \alpha\delta \mid \delta < \gamma \,\} = \alpha\gamma$.

(e). We have $0 < \omega$ and $1 < 2$, but $1\omega = \omega = 2\omega$.

(f). We show the claim $\alpha \leq \beta \to \alpha\gamma \leq \beta\gamma$ by induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned}
&\alpha\gamma \leq \beta\gamma && \text{by IH,}\\
&(\alpha\gamma) + \alpha \leq (\beta\gamma) + \alpha \leq (\beta\gamma) + \beta && \text{by Lemma 6.1(f) and (d)}\\
&\alpha(\gamma + 1) \leq \beta(\gamma + 1) && \text{by definition.}
\end{aligned}$$

Case γ limit. Then

$$\begin{aligned}
&\alpha\delta \leq \beta\delta && \text{for all } \delta < \gamma, \text{ by IH,}\\
&\alpha\delta \leq \sup\{\, \beta\delta \mid \delta < \gamma \,\},\\
&\sup\{\, \alpha\delta \mid \delta < \gamma \,\} \leq \sup\{\, \beta\delta \mid \delta < \gamma \,\},\\
&\alpha\gamma \leq \beta\gamma && \text{by definition.}
\end{aligned}$$

(g). Let $0 < \alpha$ and β limit. For the proof of $\alpha\beta$ limit we again use the characterization of limits in Lemma 3.27(a). $\alpha\beta \neq 0$: Because of $1 \leq \alpha$ and $\omega \leq \beta$ we have $0 < \omega = 1\omega \leq \alpha\beta$ by (f). $\gamma < \alpha\beta \to \gamma + 1 < \alpha\beta$: Let $\gamma < \alpha\beta = \sup\{\, \alpha\delta \mid \delta < \beta \,\}$, hence $\gamma < \alpha\delta$ for some $\delta < \beta$, hence $\gamma + 1 < \alpha\delta + 1 \leq \alpha\delta + \alpha = \alpha(\delta + 1)$ with $\delta + 1 < \beta$, hence $\gamma + 1 < \sup\{\, \alpha\delta \mid \delta < \beta \,\}$.

(h). We must show $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$. We may assume let $0 < \alpha$. We emply induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned}
\alpha[\beta + (\gamma + 1)] &= \alpha[(\beta + \gamma) + 1]\\
&= \alpha(\beta + \gamma) + \alpha\\
&= (\alpha\beta + \alpha\gamma) + \alpha && \text{by IH}\\
&= \alpha\beta + (\alpha\gamma + \alpha)\\
&= \alpha\beta + \alpha(\gamma + 1).
\end{aligned}$$

Case γ limit. By (g) $\alpha\gamma$ is a limit as well. We obtain

$$\begin{aligned}
\alpha(\beta + \gamma) &= \sup\{\, \alpha\delta \mid \delta < \beta + \gamma \,\}\\
&= \sup\{\, \alpha(\beta + \varepsilon) \mid \varepsilon < \gamma \,\}\\
&= \sup\{\, \alpha\beta + \alpha\varepsilon \mid \varepsilon < \gamma \,\} && \text{by IH}\\
&= \sup\{\, \alpha\beta + \delta \mid \delta < \alpha\gamma \,\}\\
&= \alpha\beta + \alpha\gamma.
\end{aligned}$$

(i). $(1 + 1)\omega = 2\omega = \omega$, but $1\omega + 1\omega = \omega + \omega$.

(j). If $0 < \alpha, \beta$, hence $1 \leq \alpha, \beta$, then $0 < 1 \cdot 1 \leq \alpha\beta$.

(k). Induction on γ. We may assume $\beta \neq 0$. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned}
(\alpha\beta)(\gamma + 1) &= (\alpha\beta)\gamma + \alpha\beta\\
&= \alpha(\beta\gamma) + \alpha\beta && \text{by IH}\\
&= \alpha(\beta\gamma + \beta) && \text{by (h)}\\
&= \alpha[\beta(\gamma + 1)]
\end{aligned}$$

Case γ limit. By (g) $\beta\gamma$ is a limit as well. We obtain

$$\begin{aligned}(\alpha\beta)\gamma &= \sup\{\,(\alpha\beta)\delta \mid \delta < \gamma\,\}\\ &= \sup\{\,\alpha(\beta\delta) \mid \delta < \gamma\,\} \quad \text{by IH}\\ &= \sup\{\,\alpha\varepsilon \mid \varepsilon < \beta\gamma\,\}\\ &= \alpha(\beta\gamma).\end{aligned}$$

(l). Existence: Let $0 < \beta$, hence $1 \leq \beta$ and hence $\alpha = 1\alpha \leq \beta\alpha$. Let γ be the least ordinal such that $\alpha \leq \beta\gamma$. **Case** $\alpha = \beta\gamma$. Let $\rho = 0$. **Case** $\alpha < \beta\gamma$. If $\gamma = \gamma' + 1$, then $\beta\gamma' < \alpha$. Hence there is a ρ such that $\beta\gamma' + \rho = \alpha$. Moreover, $\rho < \beta$, because from $\rho \geq \beta$ is follows that $\alpha = \beta\gamma' + \rho \geq \beta\gamma' + \beta = \beta(\gamma' + 1) = \beta\gamma$, contradicting our assumption. If γ is a limit, then $\alpha < \beta\gamma = \sup\{\,\beta\delta \mid \delta < \gamma\,\}$, hence $\alpha < \beta\delta$ for some $\delta < \gamma$, a contradiction.

Uniqueness: Assume $\beta\gamma_1 + \rho_1 = \beta\gamma_2 + \rho_2$ with $\rho_1, \rho_2 < \beta$. If say $\gamma_1 < \gamma_2$, then

$$\begin{aligned}\beta\gamma_1 + \rho_1 &< \beta\gamma_1 + \beta\\ &= \beta(\gamma_1 + 1)\\ &\leq \beta\gamma_2\\ &\leq \beta\gamma_2 + \rho_2\end{aligned}$$

hence we have a contradiction. Therefore $\gamma_1 = \gamma_2$, and hence $\rho_1 = \rho_2$. □

Corollary 6.3. *Every ordinal α can be written uniquely in the form $\alpha = \omega\gamma + n$. Here $n = 0$ iff $\alpha = 0$ or α is a limit.*

Proof. It remains to be shown that for every γ either $\omega\gamma = 0$ or $\omega\gamma$ is a limit. In case $\gamma = 0$ this is clear. In case $\gamma+1$, the ordinal $\omega(\gamma+1) = \omega\gamma+\omega$ is a limit by Lemma 6.1(h). If γ is a limit, then so is $\omega\gamma$ (by Lemma 6.2(g)). □

6.3. Ordinal Exponentiation. Ordinal exponentiation is defined by

$$\begin{aligned}\alpha^0 &:= \begin{cases}0, & \text{if } \alpha = 0;\\ 1, & \text{otherwise,}\end{cases}\\ \alpha^{\beta+1} &:= \alpha^\beta\alpha,\\ \alpha^\beta &:= \sup\{\,\alpha^\gamma \mid \gamma < \beta\,\} \quad \text{if } \beta \text{ limit.}\end{aligned}$$

Lemma 6.4 (Properties of Ordinal Exponentiation). (a) $\alpha^\beta \in \mathsf{On}$.
(b) $0^\beta = 0$, $1^\beta = \beta$.
(c) $1 < \alpha \to \beta < \gamma \to \alpha^\beta < \alpha^\gamma$.
(d) *There are α, β, γ such that $1 < \gamma$ and $1 < \alpha < \beta$, but $\alpha^\gamma \not< \beta^\gamma$.*
(e) $\alpha \leq \beta \to \alpha^\gamma \leq \beta^\gamma$.
(f) *If $1 < \alpha$ and β is a limit, then so is α^β.*
(g) $\alpha^{\beta+\gamma} = \alpha^\beta\alpha^\gamma$.
(h) $\alpha^{\beta\gamma} = (\alpha^\beta)^\gamma$.
(i) $1 < \alpha \to \beta \leq \alpha^\beta$.

Proof. (a). Induction on β. **Case** 0. Clear. **Case** $\beta + 1$. Then $\alpha^{\beta+1} = (\alpha^\beta)\alpha \in \mathsf{On}$, for by IH $\alpha^\beta \in \mathsf{On}$. **Case** β limit. Then $\alpha^\beta = \sup\{\,\alpha^\gamma \mid \gamma < \beta\,\} \in \mathsf{On}$, for by IH $\alpha^\gamma \in \mathsf{On}$ for all $\gamma < \beta$.

(b). $0^\beta = 0$: Induction on β. **Case** 0. $0^0 = 0$ holds by definition. **Case** $\beta + 1$. Then $0^{\beta+1} = (0^\beta)0 = 0$. **Case** β limit. $0^\beta = \sup\{\, 0^\gamma \mid \gamma < \beta \,\} = 0$ by IH. – $1^\beta = 1$: Induction on β. **Case** 0. Clear. **Case** $\beta + 1$. Then $1^{\beta+1} = (1^\beta)1 = 1$ by IH. **Case** β limit. $1^\beta = \sup\{\, 1^\gamma \mid \gamma < \beta \,\} = \sup\{\, 1 \mid \gamma < \beta \,\} = 1$ by IH.

(c). Let $1 < \alpha$. We show $\beta < \gamma \to \alpha^\beta < \alpha^\gamma$ by induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned}
&\beta < \gamma + 1,\\
&\beta < \gamma \vee \beta = \gamma,\\
&\alpha^\beta < \alpha^\gamma \vee \alpha^\beta = \alpha^\gamma && \text{by IH},\\
&\alpha^\beta \le \alpha^\gamma < \alpha^\gamma + \alpha^\gamma \le \alpha^{\gamma+1}.
\end{aligned}$$

Case γ limit. Let $\beta < \gamma$, hence $\beta < \delta$ for some $\delta < \gamma$. Then $\alpha^\beta < \alpha^\delta$ by IH, hence $\alpha^\beta < \sup\{\, \alpha^\delta \mid \delta < \gamma \,\} = \alpha^\gamma$.

(d). For $1 < n$ we have $n^\omega = \sup\{\, n^m \mid m < \omega \,\} = \omega$ and hence $2^\omega = \omega = 3^\omega$.

(e). We show the claim $\alpha \le \beta \to \alpha^\gamma \le \beta^\gamma$ by induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Let $\alpha \le \beta$. Then

$$\begin{aligned}
\alpha^\gamma &\le \beta^\gamma && \text{by IH},\\
\alpha^{\gamma+1} &= \alpha^\gamma \alpha\\
&\le \beta^\gamma \alpha\\
&\le \beta^\gamma \beta\\
&= \beta^{\gamma+1}.
\end{aligned}$$

Case γ limit. Let again $\alpha \le \beta$. Then

$$\begin{aligned}
&\alpha^\delta \le \beta^\delta && \text{for all } \delta < \gamma, \text{ by IH},\\
&\alpha^\delta \le \sup\{\, \beta^\delta \mid \delta < \gamma \,\}\\
&\sup\{\, \alpha^\delta \mid \delta < \gamma \,\} \le \sup\{\, \beta^\delta \mid \delta < \gamma \,\}\\
&\alpha^\gamma \le \beta^\gamma && \text{by definition}.
\end{aligned}$$

(f). Let $1 < \alpha$ and β limit. For the proof of α^β limit we again use the characterization of limits in Lemma 3.27(a). $\alpha^\beta \neq 0$: Because of $1 \le \alpha$ we have $1 = 1^\beta \le \alpha^\beta$. $\gamma < \alpha^\beta \to \gamma + 1 < \alpha^\beta$: Let $\gamma < \alpha^\beta = \sup\{\, \alpha^\delta \mid \delta < \beta \,\}$, hence $\gamma < \alpha^\delta$ for some $\delta < \beta$, hence $\gamma + 1 < \alpha^\delta + 1 \le \alpha^\delta 2 \le \alpha^{\delta+1}$ with $\delta + 1 < \beta$, hence $\gamma + 1 < \sup\{\, \alpha^\delta \mid \delta < \beta \,\}$.

(g). We must show $\alpha^{\beta+\gamma} = \alpha^\beta \alpha^\gamma$. We may assume $\alpha \neq 0, 1$. The proof is by induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned}
\alpha^{\beta+\gamma+1} &= \alpha^\beta \alpha^\gamma \alpha && \text{by IH}\\
&= \alpha^\beta \alpha^{\gamma+1}.
\end{aligned}$$

Case γ limit.

$$\begin{aligned}
\alpha^{\beta+\gamma} &= \sup\{\, \alpha^\delta \mid \delta < \beta + \gamma \,\}\\
&= \sup\{\, \alpha^{\beta+\varepsilon} \mid \varepsilon < \gamma \,\}\\
&= \sup\{\, \alpha^\beta \alpha^\varepsilon \mid \varepsilon < \gamma \,\} && \text{by IH}
\end{aligned}$$

$$\begin{aligned} &= \sup\{\, \alpha^\beta \delta \mid \delta < \alpha^\gamma \,\} \\ &= \alpha^\beta \alpha^\gamma. \end{aligned}$$

(h). We must show $\alpha^{\beta\gamma} = (\alpha^\beta)^\gamma$. We may assume $\alpha \neq 0, 1$ and $\beta \neq 0$. The proof is by induction on γ. **Case** 0. Clear. **Case** $\gamma + 1$. Then

$$\begin{aligned} \alpha^{\beta(\gamma+1)} &= \alpha^{\beta\gamma} \alpha^\beta \\ &= (\alpha^\beta)^\gamma \alpha^\beta \quad \text{by IH} \\ &= (\alpha^\beta)^{\gamma+1}. \end{aligned}$$

Case γ limit. Because of $\alpha \neq 0, 1$ and $\beta \neq 0$ we know that $\alpha^{\beta\gamma}$ and $(\alpha^\beta)^\gamma$ are limits. Hence

$$\begin{aligned} \alpha^{\beta\gamma} &= \sup\{\, \alpha^\delta \mid \delta < \beta\gamma \,\} \\ &= \sup\{\, \alpha^{\beta\varepsilon} \mid \varepsilon < \gamma \,\} \\ &= \sup\{\, (\alpha^\beta)^\varepsilon \mid \varepsilon < \gamma \,\} \quad \text{by IH} \\ &= (\alpha^\beta)^\gamma. \end{aligned}$$

(i). Let $1 < \alpha$. We show $\beta \leq \alpha^\beta$ by induction on β. **Case** 0. Clear. **Case** $\beta + 1$. Then $\beta \leq \alpha^\beta$ by IH, hence

$$\begin{aligned} \beta + 1 &\leq \alpha^\beta + 1 \\ &\leq \alpha^\beta + \alpha^\beta \\ &\leq \alpha^{\beta+1}. \end{aligned}$$

Case β limit.

$$\begin{aligned} \beta &= \sup\{\, \gamma \mid \gamma < \beta \,\} \\ &\leq \sup\{\, \alpha^\gamma \mid \gamma < \beta \,\} \quad \text{by IH} \\ &= \alpha^\beta. \end{aligned}$$

This concludes the proof. □

6.4. Cantor Normal Form.

THEOREM 6.5 (Cantor Normal Form). *Let $\gamma \geq 2$. Every α can be written uniquely in the form*

$$\alpha = \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n \quad \textit{where } \alpha \geq \alpha_1 > \cdots > \alpha_n \textit{ and } 0 < \beta_i < \gamma.$$

PROOF. Existence. Induction on α. Let δ be minimal such that $\alpha < \gamma^\delta$; such a δ exists since $\alpha \leq \gamma^\alpha$. But δ cannot be a limit, for otherwise $\alpha < \gamma^\varepsilon$ for some $\varepsilon < \delta$. If $\delta = 0$, then $\alpha = 0$ and the claim is trivial. So let $\delta = \alpha_1 + 1$, hence

$$\gamma^{\alpha_1} \leq \alpha < \gamma^{\alpha_1+1}.$$

Division with remainder gives

$$\alpha = \gamma^{\alpha_1}\beta_1 + \rho \quad \text{with } \rho < \gamma^{\alpha_1}.$$

Clearly $0 < \beta_1 < \gamma$. Now if $\rho = 0$ we are done. Otherwise we have

$$\rho = \gamma^{\alpha_2}\beta_2 + \cdots + \gamma^{\alpha_n}\beta_n \quad \text{by IH.}$$

We still must show $\alpha_1 > \alpha_2$. But this holds, because $\alpha_2 \geq \alpha_1$ entails $\rho \geq \gamma^{\alpha_2} \geq \gamma^{\alpha_1}$, a contradiction.

Uniqueness. Let

$$\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n = \gamma^{\alpha'_1}\beta'_1 + \cdots + \gamma^{\alpha'_m}\beta'_m.$$

and assume that both representations are different. Since no such sum can extend the other, we must have $i \leq n, m$ such that $(\alpha_i, \beta_i) \neq (\alpha'_i, \beta'_i)$. By Lemma 6.1(d) we can assume $i = 1$. First we have

$$\begin{aligned}
&\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}\beta_{n-1} + \gamma^{\alpha_n}\beta_n \\
&< \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}\beta_{n-1} + \gamma^{\alpha_n+1} && \text{since } \beta_n < \gamma \\
&\leq \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}}(\beta_{n-1}+1) && \text{for } \alpha_n < \alpha_{n-1} \\
&\leq \gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_{n-1}+1} \\
&\cdots \\
&\leq \gamma^{\alpha_1}(\beta_1 + 1).
\end{aligned}$$

Now if e.g. $\alpha_1 < \alpha'_1$, then we would have $\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n < \gamma^{\alpha_1}(\beta_1+1) \leq \gamma^{\alpha_1+1} \leq \gamma^{\alpha'_1}$, which cannot be. Hence $\alpha_1 = \alpha'_1$. If e.g. $\beta_1 < \beta'_1$, then we would have $\gamma^{\alpha_1}\beta_1 + \cdots + \gamma^{\alpha_n}\beta_n < \gamma^{\alpha_1}(\beta_1 + 1) \leq \gamma^{\alpha_1}\beta'_1$, which again cannot be the case. Hence $\beta_1 = \beta'_1$. □

Corollary 6.6 (Cantor Normal Form With Base ω). *Every α can be written uniquely in the form*

$$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n} \quad \textit{with } \alpha \geq \alpha_1 \geq \cdots \geq \alpha_n.$$

An ordinal α is an *additive principal number* when $\alpha \neq 0$ and $\beta + \gamma < \alpha$ for $\beta, \gamma < \alpha$.

Corollary 6.7. *Additive principal numbers are exactly the ordinals of the form ω^ξ.*

Proof. This follows from Cantor's normal form with base ω. □

Corollary 6.8 (Cantor Normal Form With Base 2). *Every α can be written uniquely in the form*

$$\alpha = 2^{\alpha_1} + \cdots + 2^{\alpha_n} \quad \textit{with } \alpha \geq \alpha_1 > \cdots > \alpha_n.$$

Let $\omega_0 := 1$, $\omega_{k+1} := \omega^{\omega_k}$ and $\varepsilon_0 := \sup_{k<\omega} \omega_k$. Notice that ε_0 is the least ordinal α such that $\omega^\alpha = \alpha$.

7. Normal Functions

In [**29**] Veblen investigated the notion of a continuous monotonic function on a segment of the ordinals, and introduced a certain hierarchy of normal functions. His goal was to generalize Cantor's theory of ε-numbers (from [**5**]).

7.1. Closed Unbounded Classes. Let Ω be a regular cardinal $> \omega$ or $\Omega = \mathsf{On}$. An important example is $\Omega = \aleph_1$, that is the case where Ω is the set of all countable ordinals. Let $\alpha, \beta, \gamma, \delta, \varepsilon, \xi, \eta, \zeta$ denote elements of Ω. A function $\varphi\colon \Omega \to \Omega$ is *monotone* if $\alpha < \beta$ implies $\varphi\alpha < \varphi\beta$. φ is *continuous* if $\varphi\alpha = \sup_{\xi<\alpha} \varphi\xi$ for every limit α. φ is *normal* if φ is monotone and continuous.

LEMMA 7.1. *For every monotone function φ we have $\alpha \leq \varphi\alpha$.*

PROOF. Induction on α. **Case** 0. $0 \leq \varphi 0$. **Case** $\alpha + 1$. $\alpha \leq \varphi\alpha < \varphi(\alpha + 1)$. **Case** α limit. $\alpha = \sup_{\xi<\alpha} \xi \leq \sup_{\xi<\alpha} \varphi\xi \leq \varphi\alpha$. □

A class $\mathcal{B} \subseteq \Omega$ is *bounded* if $\sup(\mathcal{B}) \in \Omega$. A class $\mathcal{A} \subseteq \Omega$ is *closed* if for every bounded subclass $\mathcal{B} \subseteq \mathcal{A}$ we have $\sup(\mathcal{B}) \in \mathcal{A}$. Closed unbounded classes $\mathcal{A} \subseteq \Omega$ are called *normal* or *closed unbounded* in Ω (club for short).

If for instance $\Omega = \Omega_1$, then every $\mathcal{B} \subseteq \Omega$ is a set, and $\mathcal{B}$ is bounded iff $\mathcal{B}$ is countable. If $\Omega = \mathsf{On}$, then $\mathcal{B}$ is bounded iff $\mathcal{B}$ is a set.

By Corollary 3.19 (to the Isomorphy Theorem of Mostowski) for every $\mathcal{A} \subseteq \mathsf{On}$ we have a uniquely determined isomorphism of an ordinal class onto $\mathcal{A}$, that is an $f\colon \mathsf{On} \to \mathcal{A}$ (or $f\colon \alpha \to \mathcal{A}$). This isomorphism is called the *ordering function* of $\mathcal{A}$. Notice that f is the *monotone enumeration* of $\mathcal{A}$.

LEMMA 7.2. *The range of a normal function is a normal class. Conversely, the ordering function of a normal class is a normal function.*

PROOF. Let φ be a normal function. $\varphi[\Omega]$ is unbounded, since for every α we have $\alpha \leq \varphi\alpha$. We now show that $\varphi[\Omega]$ is closed. So let $\mathcal{B} = \{\, \varphi\xi \mid \xi \in \mathcal{A} \,\}$ be bounded, i.e., $\sup(\mathcal{B}) \in \Omega$. Because of $\xi \leq \varphi\xi$ then also $\mathcal{A}$ is bounded. We must show $\sup(\mathcal{B}) = \varphi\alpha$ for some α. If $\mathcal{A}$ has a maximal element we are done. Otherwise $\alpha := \sup(\mathcal{A})$ is a limit. Then $\varphi\alpha = \sup_{\xi<\alpha} \varphi\xi = \sup_{\xi\in\mathcal{A}} \varphi\xi = \sup(\mathcal{B})$. Conversely, let $\mathcal{A}$ be closed and unbounded. We define a function $\varphi\colon \Omega \to \mathcal{A}$ by transfinite recursion, as follows.

$$\varphi\alpha := \min\{\, \gamma \in \mathcal{A} \mid \forall\xi.\xi < \alpha \to \varphi\xi < \gamma \,\}.$$

φ is well defined, since $\mathcal{A}$ is unbounded. Clearly φ is the ordering function of $\mathcal{A}$ and hence monotone. It remains to be shown that φ is continuous. So let α be a limit. Since $\varphi[\alpha]$ is bounded (this follows from $\varphi\xi < \varphi\alpha$ for $\xi < \alpha$) and $\mathcal{A}$ is closed, we have $\sup_{\xi\in\alpha} \varphi\xi \in \mathcal{A}$, hence by definition $\varphi\alpha = \sup_{\xi\in\alpha} \varphi\xi$. □

LEMMA 7.3. *The fixed points of a normal function form a normal class.*

PROOF. (Cf. Cantor [**5**, p. 242]). Let φ be a normal function. For every ordinal α we can construct a fixed point $\beta \geq \alpha$ of φ by

$$\beta := \sup\{\, \varphi^n\alpha \mid n \in \mathbb{N} \,\}.$$

Hence the class of fixed points of φ is unbounded. It is closed as well, since for every class $\mathcal{B}$ of fixed points of φ we have $\varphi(\sup(\mathcal{B})) = \sup\{\, \varphi\alpha \mid \alpha \in \mathcal{B} \,\} = \sup\{\, \alpha \mid \alpha \in \mathcal{B} \,\} = \sup(\mathcal{B})$, i.e., $\sup(\mathcal{B})$ is a fixed point of φ. □

7.2. The Veblen Hierarchy of Normal Functions. The ordering function of the class of fixed points of a normal function φ has been called by Veblen the *first derivative* φ' of φ. For example, the first derivative of the function ω^ξ is the function ε_ξ.

LEMMA 7.4 (Veblen). *Let $(\mathcal{A}_\gamma)_{\gamma<\beta}$ with β limit be a decreasing sequence of normal classes. Then the intersection $\bigcap_{\gamma<\beta}\mathcal{A}_\gamma$ is normal as well.*

PROOF. Unboundedness. Let α be given and $\delta_\gamma := \min\{\,\xi \in \mathcal{A}_\gamma \mid \xi > \alpha\,\}$. Then $(\delta_\gamma)_{\gamma<\beta}$ is weakly monotonic. Let $\delta := \sup_{\gamma<\beta}\delta_\gamma$. Then $\delta \in \mathcal{A}_\gamma$ for every $\gamma<\beta$, since the $\mathcal{A}_\gamma$ decrease. Hence $\alpha < \delta \in \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$.

Closedness. Let $\mathcal{B} \subseteq \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$, $\mathcal{B}$ bounded. Then $\mathcal{B} \subseteq \mathcal{A}_\gamma$ for every $\gamma<\beta$ and therefore $\sup(\mathcal{B}) \in \mathcal{A}_\gamma$. Hence $\sup(\mathcal{B}) \in \bigcap_{\gamma<\beta}\mathcal{A}_\gamma$. □

We now define the *Veblen hierarchy of normal functions*. It is based on an arbitrary given normal function $\varphi\colon \Omega \to \Omega$. We use transfinite recursion to define for every $\beta \in \Omega$ a normal function $\varphi_\beta\colon \Omega \to \Omega$:

$$\varphi_0 := \varphi,$$
$$\varphi_{\beta+1} := (\varphi_\beta)',$$
for limits β let φ_β be the ordering function of $\bigcap_{\gamma<\beta}\varphi_\gamma[\Omega]$.

For example, for $\varphi\alpha := 1+\alpha$ we obtain $\varphi_\beta\alpha = \omega^\beta + \alpha$. If we start with $\varphi\alpha := \omega^\alpha$, then $\varphi_1\alpha = \varepsilon_\alpha$ and φ_2 enumerates the critical ε-numbers, i.e., the ordinals α such that $\varepsilon_\alpha = \alpha$.

LEMMA 7.5. *Let $\beta > 0$. Then φ_β is the ordering function of the class of all common fixed points of all φ_γ for $\gamma<\beta$.*

PROOF. We must show $\varphi_\beta[\Omega] = \{\,\xi \mid \forall\gamma.\gamma<\beta \to \varphi_\gamma\xi = \xi\,\}$.

$\subseteq$. This is proved by transfinite induction on β. In case $\beta+1$ every $\varphi_{\beta+1}\alpha$ is a fixed point of φ_β and hence by IH also a fixed point of all φ_γ for $\gamma<\beta$. If β is a limit, then the claim follows from $\varphi_\beta[\Omega] = \bigcap_{\gamma<\beta}\varphi_\gamma[\Omega]$.

$\supseteq$. Let ξ such that $\forall\gamma.\gamma<\beta \to \varphi_\gamma\xi = \xi$ be given. If β is a successor, then $\xi \in \varphi_\beta[\Omega]$ by definition of φ_β. If β is a limit, then $\xi \in \bigcap_{\gamma<\beta}\varphi_\gamma[\Omega] = \varphi_\beta[\Omega]$. □

It follows that $\varphi_\gamma(\varphi_\beta\xi) = \varphi_\beta\xi$ for every $\gamma<\beta$.

A further normal function can be obtained as follows. From each of the normal classes $\varphi_\beta[\Omega]$ pick the least fixed point. The class formed in this way again is normal, hence can be enumerated by a normal function. This normal function assigns to every β the ordinal $\varphi_\beta 0$.

LEMMA 7.6. *If φ is a normal function with $0 < \varphi 0$, then $\lambda\beta\,\varphi_\beta 0$ is a normal function as well.*

PROOF. We first show

$$\beta<\gamma \to \varphi_\beta 0 < \varphi_\gamma 0,$$

by induction on γ. So let $\beta<\gamma$. Observe that $0 < \varphi_\beta 0$ by IH or in case $\beta = 0$ by assumption. Hence 0 is not a fixed point of φ_β and therefore $0 < \varphi_\gamma 0$. But this implies $\varphi_\beta 0 < \varphi_\beta(\varphi_\gamma 0) = \varphi_\gamma 0$.

We now show that $\lambda\beta\varphi_\beta 0$ is continuous. Let $\delta := \sup_{\beta<\gamma}\varphi_\beta 0$ with γ limit. We must show $\delta = \varphi_\gamma 0$. Because of $\varphi_\beta 0 \in \varphi_\alpha[\Omega]$ for all $\alpha \le \beta < \gamma$

and since $\varphi_\alpha[\Omega]$ is closed we have $\delta \in \varphi_\alpha[\Omega]$, hence $\delta \in \bigcap_{\alpha<\gamma} \varphi_\alpha[\Omega] = \varphi_\gamma[\Omega]$ and therefore $\delta \geq \varphi_\gamma 0$. On the other hand $\varphi_\beta 0 < \varphi_\beta(\varphi_\gamma 0) = \varphi_\gamma 0$, hence $\delta \leq \varphi_\gamma 0$. □

The fixed points of this function, i.e., the ordinals α such that $\varphi_\alpha 0 = \alpha$, are called *strongly critical* ordinals. Observe that they depend on the given normal function $\varphi = \varphi_0$. Their ordering function is usually denoted by Γ. Hence by definition $\Gamma_0 := \Gamma 0$ is the least ordinal β such that $\varphi_\beta 0 = \beta$.

7.3. φ Normal Form. We now generalize Cantor's normal form, using the Veblen hierarchy instead of ω^ξ.

Lemma 7.7.

$$\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1 \iff \begin{cases} \alpha_0 < \varphi_{\beta_1}\alpha_1, & \text{if } \beta_0 < \beta_1; \\ \alpha_0 < \alpha_1, & \text{if } \beta_0 = \beta_1; \\ \varphi_{\beta_0}\alpha_0 < \alpha_1, & \text{if } \beta_0 > \beta_1, \end{cases} \tag{41}$$

$$\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\alpha_1 \iff \begin{cases} \alpha_0 = \varphi_{\beta_1}\alpha_1, & \text{if } \beta_0 < \beta_1; \\ \alpha_0 = \alpha_1, & \text{if } \beta_0 = \beta_1; \\ \varphi_{\beta_0}\alpha_0 = \alpha_1, & \text{if } \beta_0 > \beta_1. \end{cases} \tag{42}$$

Proof. $\Leftarrow$. (41). If $\beta_0 < \beta_1$ and $\alpha_0 < \varphi_{\beta_1}\alpha_1$, then $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_0}\varphi_{\beta_1}\alpha_1 = \varphi_{\beta_1}\alpha_1$. If $\beta_0 = \beta_1$ and $\alpha_0 < \alpha_1$, then $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$. If $\beta_0 > \beta_1$ and $\varphi_{\beta_0}\alpha_0 < \alpha_1$, then $\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$. For (42) one argues similarly.

$\Rightarrow$. If the right hand side of (41) is false, we have

$$\begin{cases} \alpha_1 \leq \varphi_{\beta_0}\alpha_0, & \text{if } \beta_1 < \beta_0; \\ \alpha_1 \leq \alpha_0, & \text{if } \beta_1 = \beta_0; \\ \varphi_{\beta_1}\alpha_1 \leq \alpha_0, & \text{if } \beta_1 > \beta_0, \end{cases}$$

hence by $\Leftarrow$ (with 0 and 1 exchanged) $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta_0}\alpha_0$ or $\varphi_{\beta_1}\alpha_1 = \varphi_{\beta_0}\alpha_0$, hence $\neg(\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1)$. If the right hand side of (42) is false, we have

$$\begin{cases} \alpha_0 \neq \varphi_{\beta_1}\alpha_1, & \text{if } \beta_0 < \beta_1; \\ \alpha_0 \neq \alpha_1, & \text{if } \beta_0 = \beta_1; \\ \varphi_{\beta_0}\alpha_0 \neq \alpha_1, & \text{if } \beta_0 > \beta_1, \end{cases}$$

and hence by $\Leftarrow$ in (41) either $\varphi_{\beta_0}\alpha_0 < \varphi_{\beta_1}\alpha_1$ or $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta_0}\alpha_0$, hence $\varphi_{\beta_0}\alpha_0 \neq \varphi_{\beta_1}\alpha_1$. □

Corollary 7.8. *If $\beta_0 \leq \beta_1$, then $\varphi_{\beta_0}\alpha \leq \varphi_{\beta_1}\alpha$.*

Proof. Assume $\beta_0 < \beta_1$. By Lemma 7.7 (for $\leq$) it suffices to show $\alpha \leq \varphi_{\beta_1}\alpha$. But this follows from Lemma 7.1. □

Corollary 7.9. *If $\varphi_{\beta_0}\alpha_0 = \varphi_{\beta_1}\alpha_1$, then $\alpha_0 = \alpha_1$ and $\beta_0 = \beta_1$, provided $\alpha_0 < \varphi_{\beta_0}\alpha_0$ and $\alpha_1 < \varphi_{\beta_1}\alpha_1$.*

Proof. **Case** $\beta_0 = \beta_1$. Then $\alpha_0 = \alpha_1$ follows from Lemma 7.7. **Case** $\beta_0 < \beta_1$. By Lemma 7.7 we have $\alpha_0 = \varphi_{\beta_1}\alpha_1 = \varphi_{\beta_0}\alpha_0$, contradicting our assumption. **Case** $\beta_1 < \beta_0$. Similar. □

COROLLARY 7.10. *If φ is a normal function with $0 < \varphi 0$, then every fixed point α of $\varphi = \varphi_0$ can be written uniquely in the form $\alpha = \varphi_\beta \alpha'$ with $\alpha' < \alpha$.*

PROOF. We have $\alpha + 1 \leq \varphi_{\alpha+1} 0$ by Lemma 7.6 and hence $\alpha < \varphi_{\alpha+1}\alpha$. Now let β be minimal such that $\alpha < \varphi_\beta \alpha$. By assumption $0 < \beta$. Since α is a fixed point of all φ_γ with $\gamma < \beta$, we have $\alpha = \varphi_\beta \alpha'$ for some α'. Using $\alpha < \varphi_\beta \alpha$ it follows that $\alpha' < \alpha$.

Uniqueness. Let in addition $\alpha = \varphi_{\beta_1}\alpha_1$ with $\alpha_1 < \alpha$. Then $\alpha_1 < \varphi_{\beta_1}\alpha_1$, hence $\beta \leq \beta_1$ by the choice of β. Now if $\beta < \beta_1$, then we would obtain $\varphi_\beta \alpha = \varphi_\beta \varphi_{\beta_1}\alpha_1 = \varphi_{\beta_1}\alpha_1 = \alpha$, contradicting our choice of β. Hence $\beta = \beta_1$ and therefore $\alpha_1 = \alpha$. □

We now show that every ordinal can be written uniquely in a certain φ normal form. Here we assume that our initial normal function $\varphi_0 = \varphi$ is the exponential function with base ω.

THEOREM 7.11 (φ Normal Form). *Let $\varphi_0 \xi := \omega^\xi$. Then every ordinal α can be written uniquely in the form*

$$\alpha = \varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n$$

with $\varphi_{\beta_1}\alpha_1 \geq \cdots \geq \varphi_{\beta_n}\alpha_n$ and $\alpha_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, k$. If $\alpha < \Gamma_0$, then in addition we have $\beta_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, n$.

PROOF. Existence. First write α in Cantor normal form $\alpha = \varphi_0\delta_1 + \cdots + \varphi_0\delta_n$ with $\delta_1 \geq \cdots \geq \delta_n$. Every summand with $\delta_i < \varphi_0\delta_i$ is left unchanged. Every other summand satisfies $\delta_i = \varphi_0\delta_i$ and hence by Corollary 7.10 can be replaced by $\varphi_\beta \alpha'$ where $\alpha' < \varphi_\beta \alpha'$.

Uniqueness. Let

$$\alpha = \varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n = \varphi_{\beta'_1}\alpha'_1 + \cdots + \varphi_{\beta'_m}\alpha'_m$$

and assume that both representations are different. Since no such sum can extend the other, we must have $i \leq n, m$ such that $(\beta_i, \alpha_i) \neq (\beta'_i, \alpha'_i)$. By Lemma 6.1(d) we can assume $i = 1$. Now if say $\varphi_{\beta_1}\alpha_1 < \varphi_{\beta'_1}\alpha'_1$, then we would have (since $\varphi_{\beta'_1}\alpha'_1$ is an additive principal number and $\varphi_{\beta_1}\alpha_1 \geq \cdots \geq \varphi_{\beta_n}\alpha_n$)

$$\varphi_{\beta_1}\alpha_1 + \cdots + \varphi_{\beta_n}\alpha_n < \varphi_{\beta'_1}\alpha'_1 \leq \varphi_{\beta'_1}\alpha'_1 + \cdots + \varphi_{\beta'_m}\alpha'_m,$$

a contradiction.

We must show that in case $\alpha < \Gamma_0$ we have $\beta_i < \varphi_{\beta_i}\alpha_i$ for $i = 1, \ldots, n$. So assume $\varphi_{\beta_i}\alpha_i \leq \beta_i$ for some i. Then

$$\varphi_{\beta_i} 0 \leq \varphi_{\beta_i}\alpha_i \leq \beta_i \leq \varphi_{\beta_i} 0,$$

hence $\varphi_{\beta_i} 0 = \beta_i$ and hence

$$\Gamma_0 \leq \beta_i = \varphi_{\beta_i} 0 \leq \varphi_{\beta_i}\alpha_i \leq \alpha.$$

□

From the $\varphi_\beta(\alpha)$ one obtains a unique notation system for ordinals below $\Gamma_0 := \Gamma 0$. Observe however that $\Gamma_0 = \varphi_{\Gamma_0} 0$. by definition of Γ_0.

8. Notes

Set theory as presented in these notes is commonly called ZFC (Zermelo-Fraenkel set theory with the axiom of choice; C for Choice). Zermelo wrote these axioms in 1908, with the exeptions of the Regularity Axioms (of von Neumann, 1925) and the replacement scheme (Fraenkel, 1922). Also Skolem considered principles related to these additional axioms. In ZFC the only objects are sets, classes are only a convenient way to speak of formulas.

The hierarchy of normal functions defined in Section 7 has been extended by Veblen [**29**] to functions with more than one argument. Schütte in [**21**] studied these functions carefully and could show that they can be used for a cosntructive representation of a segment of the ordinals far bigger than Γ_0. To this end he introduced so-called inquotesKlammersymbole to denote the multiary Veblen-functions.

Bachmann extended the Veblen hierarchy using the first uncountable ordinal Ω. His approach has later been extended by means of symbols for higher number classes, first by Pfeiffer for finite number clasees and then by Isles for transfinite number classes. However, the resulting theory was quite complicated and difficult to work with. An idea of Feferman then simplified the subject considerably. He introduced functions $\theta_\alpha\colon \mathsf{On} \to \mathsf{On}$ for $\alpha \in \mathsf{On}$, that form again a hierarchy of normal functions and extend the Veblen hierarchy. One usually writes $\theta\alpha\beta$ instad of $\theta_\alpha(\beta)$ and views θ as a binay function. The ordinals $\theta\alpha\beta$ can be defined byransfinite recursion on α, as follows. Assume thatd θ_ξ for every $\xi < \alpha$ is defined already. Let $C(\alpha, \beta)$ be the set of all ordinals that can be generated from ordinals $< \beta$ and say the constants $0, \aleph_1, \ldots, \aleph_\omega$ by means of the functions $+$ and $\theta{\restriction}\{\, \xi \mid \xi < \alpha \,\} \times \mathsf{On}$. An ordinal β is aclled α-critical if $\beta \notin C(\alpha, \beta)$. Than $\theta_\alpha\colon \mathsf{On} \to \mathsf{On}$ is defined as the ordering function of the class of all α-critical ordinals.

Buchholz observed in [**4**] that the second argument β in $\theta\alpha\beta$ is not used in any essential way, and that the functions $\alpha \mapsto \theta\alpha\aleph_\nu$ with $\nu = 0, 1, \ldots, \omega$ generate a notation system for ordinals of the same strength as the system with the binary θ-function. He then went on and defined directly functions ψ_ν with $\nu \leq \omega$, that correspond to $\alpha \mapsto \theta\alpha\aleph_\nu$. More precisely he defined $\psi_\nu\alpha$ for $\alpha \in \mathsf{On}$ and $\nu \leq \omega$ by transfinite recursion on α (simultaneously for all ν), as follows.

$$\psi_\nu\alpha := \min\{\, \gamma \mid \gamma \notin C_\nu(\alpha) \,\},$$

where $C_\nu(\alpha)$ is the set of all ordinals that can be generated from the ordinals $< \aleph_\nu$ by the functions $+$ and all $\psi_u{\restriction}\{\, \xi \mid \xi < \alpha \,\}$ with $u \leq \omega$.

CHAPTER 6

Proof Theory

This chapter presents an example of the type of proof theory inspired by Hilbert's programme and the Gödel incompleteness theorems. The principal goal will be to offer an example of a true mathematically meaningful principle not derivable in first-order arithmetic.

The main tool for proving theorems in arithmetic is clearly the induction schema

$$A(0) \to (\forall x.A(x) \to A(S(x))) \to \forall x A(x).$$

Here $A(x)$ is an arbitrary formula. An equivalent form of this schema is "cumulative" or course-of-values induction

$$(\forall x.\forall y{<}x A(y) \to A(x)) \to \forall x A(x).$$

Both schemes refer to the standard ordering of the natural numbers. Now it is tempting to try to strengthen arithmetic by allowing more general induction schemas, e.g. with respect to the lexicographical ordering of $\mathbb{N}\times\mathbb{N}$. More generally, we might pick an arbitrary well-ordering $\prec$ over $\mathbb{N}$ and use the schema of *transfinite induction*:

$$(\forall x.\forall y{\prec}x\, A(y) \to A(x)) \to \forall x A(x).$$

This can be read as follows. Suppose the property $A(x)$ is "progressive", i.e. from the validity of $A(y)$ for all $y \prec x$ we can always conclude that $A(x)$ holds. Then $A(x)$ holds for all x.

One might wonder whether this schema of transfinite induction actually strengthens arithmetic. We will prove here a classic result of Gentzen [**9**] which in a sense answers this question completely. However, in order to state the result we have to be more explicit about the well-orderings used. This is done in the next section.

1. Ordinals Below ε_0

In order to be able to speak in arithmetical theories about ordinals, we use use a Gödelization of ordinals. This clearly is possible for countable ordinals only. Here we restrict ourselves to a countable set of relatively small ordinals, the ordinals below ε_0. Moreover, we equip these ordinals with an extra structure (a kind of algebra). It is then customary to speak of *ordinal notations*. These ordinal notations could be introduced without any set theory in a purely formal, combinatorial way, based on the Cantor normal form for ordinals. However, we take advantage of the fact that we have just dealt with ordinals within set theory. We also introduce some elementary relations and operations for such ordinal notations, which will be used later. For brevity we from now on use the word "ordinal" instead of "ordinal notation".

1.1. Comparison of Ordinals; Natural Sum.

LEMMA 1.1. *Let $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ and $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ be Cantor normal forms (with $m,n\geq -1$). Then*

$$\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}<\omega^{\beta_n}+\cdots+\omega^{\beta_0}$$

iff there is an $i\geq 0$ such that $\alpha_{m-i}<\beta_{n-i}$, $\alpha_{m-i+1}=\beta_{n-i+1},\ldots,\alpha_m=\beta_n$, or $m<n$ and $\alpha_m=\beta_n,\ldots,\alpha_0=\beta_{n-m}$.

PROOF. Exercise. □

We use the notations 1 for ω^0, a for $\omega^0+\cdots+\omega^0$ with a copies of ω^0 and $\omega^\alpha a$ for $\omega^\alpha+\cdots+\omega^\alpha$ again with a copies of ω^α.

LEMMA 1.2. *Let $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ and $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ be Cantor normal forms. Then*

$$\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}+\omega^{\beta_n}+\cdots+\omega^{\beta_0}=\omega^{\alpha_m}+\cdots+\omega^{\alpha_i}+\omega^{\beta_n}+\cdots+\omega^{\beta_0},$$

where i is minimal such that $\alpha_i\geq\beta_n$; if there is no such i, let $i=m+1$ (so $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$).

PROOF. Exercise. □

One can also define a commutative variant of addition. This is the so-called *natural sum* or *Hessenberg sum* of two ordinals. For Cantor normal forms $\omega^{\alpha_m}+\cdots+\omega^{\alpha_0}$ and $\omega^{\beta_n}+\cdots+\omega^{\beta_0}$ it is defined by

$$(\omega^{\alpha_m}+\cdots+\omega^{\alpha_0})\#(\omega^{\beta_n}+\cdots+\omega^{\beta_0}):=\omega^{\gamma_{m+n+1}}+\cdots+\omega^{\gamma_0},$$

where $\gamma_{m+n+1},\ldots,\gamma_0$ is a decreasing permutation of $\alpha_m,\ldots,\alpha_0,\beta_n,\ldots,\beta_0$.

LEMMA 1.3. *$\#$ is associative, commutative and strongly monotonic in both arguments.*

PROOF. Exercise. □

1.2. Enumerating Ordinals. In order to work with ordinals in a purely arithmetical system we set up some effective bijection between our ordinals $<\varepsilon_0$ and non-negative integers (i.e., a Gödel numbering). For its definition it is useful to refer to ordinals in the form

$$\omega^{\alpha_m}k_m+\cdots+\omega^{\alpha_0}k_0\quad\text{with }\alpha_m>\cdots>\alpha_0\text{ and }k_i\neq 0\ (m\geq -1).$$

(By convention, $m=-1$ corresponds to the empty sum.)

For every ordinal α we define its Gödel number $\ulcorner\alpha\urcorner$ inductively by

$$\ulcorner\omega^{\alpha_m}k_m+\cdots+\omega^{\alpha_0}k_0\urcorner:=\Big(\prod_{i\leq m}p_{\ulcorner\alpha_i\urcorner}^{k_i}\Big)-1,$$

where p_n is the n-th prime number starting with $p_0:=2$. For every non-negative integer x we define its corresponding ordinal notation $\mathsf{o}(x)$ inductively by

$$\mathsf{o}\Big((\prod_{i\leq l}p_i^{q_i})-1\Big):=\sum_{i\leq l}\omega^{\mathsf{o}(i)}q_i,$$

where the sum is to be understood as the natural sum.

LEMMA 1.4. *(a)* $\mathsf{o}(\ulcorner\alpha\urcorner)=\alpha$,
(b) $\ulcorner\mathsf{o}(x)\urcorner=x$.

PROOF. This can be proved easily by induction. □

Hence we have a simple bijection between ordinals and non-negative integers. Using this bijection we can transfer our relations and operations on ordinals to computable relations and operations on non-negative integers. We use the following abbreviations.

$$\begin{array}{lcl} x \prec y & := & \mathsf{o}(x) < \mathsf{o}(y), \\ \omega^x & := & \ulcorner \omega^{\mathsf{o}(x)} \urcorner, \\ x \oplus y & := & \ulcorner \mathsf{o}(x) + \mathsf{o}(y) \urcorner, \\ xk & := & \ulcorner \mathsf{o}(x) k \urcorner, \\ \omega_k & := & \ulcorner \omega_k \urcorner, \end{array}$$

where $\omega_0 := 1$, $\omega_{k+1} := \omega^{\omega_k}$.

We leave it to the reader to verify that $\prec$, $\lambda x.\omega^x$, $\lambda xy.x \oplus y$, $\lambda xk.xk$ and $\lambda k.\ulcorner \omega_k \urcorner$ are all primitive recursive.

2. Provability of Initial Cases of TI

We now derive initial cases of the principle of transfinite induction in arithmetic, i.e., of

$$(\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x{\prec}a\, Px$$

for some number a and a predicate symbol P, where $\prec$ is the standard order of order type ε_0 defined in the preceeding section. In a later section we will see that our results here are optimal in the sense that for the full system of ordinals $< \varepsilon_0$ the principle

$$(\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x Px$$

of transfinite induction is underivable. All these results are due to Gentzen [**9**].

2.1. Arithmetical Systems. By an *arithmetical system* **Z** we mean a theory based on minimal logic in the $\forall{\to}\bot$-language (including equality axioms), with the following properties. The language of **Z** consists of a fixed (possibly countably infinite) supply of function and relation constants which are assumed to denote fixed functions and relations on the non-negative integers for which a computation procedure is known. Among the function constants there must be a constant S for the successor function and 0 for (the 0-place function) zero. Among the relation constants there must be a constant $=$ for equality and $\prec$ for the ordering of type ε_0 of the natural numbers, as introduced in Section 1. In order to formulate the general principle of transfinite induction we also assume that a unary relation symbol P is present, which acts like a free set variable.

Terms are built up from object variables x, y, z by means of $f(t_1, \dots, t_m)$, where f is a function constant. We identify closed terms which have the same value; this is a convenient way to express in our formal systems the assumption that for each function constant a computation procedure is known. Terms of the form $\mathsf{S}(\mathsf{S}(\dots \mathsf{S}(0) \dots))$ are called *numerals*. We use the notation $\mathsf{S}^n 0$ or $\overline{n}$ or (only in this chapter) even n for them. *Formulas* are built up from $\bot$ and atomic formulas $R(t_1, \dots, t_m)$, with R a relation constant or

a relation symbol, by means of $A \to B$ and $\forall x A$. Recall that we abbreviate $A \to \bot$ by $\neg A$.

The *axioms* of $\mathbf{Z}$ will always include the *Peano axioms*, i.e., the universal closures of

$$\mathsf{S}(x) = \mathsf{S}(y) \to x = y, \tag{43}$$

$$\mathsf{S}(x) = 0 \to A, \tag{44}$$

$$A(0) \to (\forall x.A(x) \to A(\mathsf{S}(x))) \to \forall x A(x), \tag{45}$$

with $A(x)$ an arbitrary formula. We express our assumption that for every relation constant R a decision procedure is known by adding the axiom $R\vec{n}$ whenever $R\vec{n}$ is true, and $\neg R\vec{n}$ whenever $R\vec{n}$ is false. Concerning $\prec$ we require irreflexivity and transitivity for $\prec$ as axioms, and also – following Schütte – the universal closures of

$$x \prec 0 \to A, \tag{46}$$

$$z \prec y \oplus \omega^0 \to (z \prec y \to A) \to (z = y \to A) \to A, \tag{47}$$

$$x \oplus 0 = x, \tag{48}$$

$$x \oplus (y \oplus z) = (x \oplus y) \oplus z, \tag{49}$$

$$0 \oplus x = x, \tag{50}$$

$$\omega^x 0 = 0, \tag{51}$$

$$\omega^x \mathsf{S}(y) = \omega^x y \oplus \omega^x, \tag{52}$$

$$z \prec y \oplus \omega^{\mathsf{S}(x)} \to z \prec y \oplus \omega^{e(x,y,z)} m(x,y,z), \tag{53}$$

$$z \prec y \oplus \omega^{\mathsf{S}(x)} \to e(x,y,z) \prec \mathsf{S}(x), \tag{54}$$

where $\oplus$, $\lambda xy.\omega^x y$, e and m denote the appropriate function constants and A is any formula. (The reader should check that e, m can be taken to be primitive recursive.) These axioms are formal counterparts to the properties of the ordinal notations observed in the preceeding section. We also allow an arbitrary supply of true formulas $\forall \vec{x} A$ with A quantifier-free and without P as axioms. Such formulas are called Π_1*-formulas* (in the literature also Π^0_1-formulas).

Moreover, we may also add an *ex-falso-quodlibet schema* Efq or even a *stability schema* Stab for A:

$$\forall x.\bot \to A,$$
$$\forall x.\neg\neg A \to A.$$

Addition of Efq leads to an intuitionistic arithmetical system (the $\forall{\to}\bot$-fragment of of Heyting arithmetic **HA**), and addition of Stab to a classical arithmetical system (a version of Peano arithmetic **PA**). Note that in our $\forall{\to}\bot$-fragment of minimal logic these schemas are derivable from their instances

$$\forall \vec{x}.\bot \to R\vec{x},$$
$$\forall \vec{x}.\neg\neg R\vec{x} \to R\vec{x},$$

with R a relation constant or the special relation symbol P. Note also that when the stability schema is present, we can replace (44), (46) and (47) by

their more familiar classical versions

$$\mathsf{S}(x) \neq 0, \tag{55}$$

$$x \not\prec 0, \tag{56}$$

$$z \prec y \oplus \omega^0 \to z \neq y \to z \prec y. \tag{57}$$

We will also consider *restricted* arithmetical systems $\mathbf{Z}_k$. They are defined like $\mathbf{Z}$, but with the induction schema (45) restricted to formulas A of level $\mathsf{lev}(A) \leq k$. The *level* of a formula A is defined by

$$\begin{aligned} \mathsf{lev}(R\vec{t}) &:= \mathsf{lev}(\bot) := 0, \\ \mathsf{lev}(A \to B) &:= \max(\mathsf{lev}(A) + 1, \mathsf{lev}(B)), \\ \mathsf{lev}(\forall x A) &:= \max(1, \mathsf{lev}(A)). \end{aligned}$$

However, the trivial special case of induction $A(0) \to \forall x A(\mathsf{S}x) \to \forall x A$, which amounts to case distinction, is allowed for arbitrary A. (This is needed in the proof of Theorem 2.2 below)

2.2. Gentzen's Proof.

THEOREM 2.1 (Provable Initial Cases of TI in $\mathbf{Z}$). *Transfinite induction up to ω_n, i.e., for arbitrary $A(x)$ the formula*

$$(\forall x.\forall y{\prec}x\, A(y) \to A(x)) \to \forall x{\prec}\omega_n\, A(x),$$

is derivable in $\mathbf{Z}$.

PROOF. To every formula $A(x)$ we assign a formula $A^+(x)$ (with respect to a fixed variable x) by

$$A^+(x) := (\forall y.\forall z{\prec}y\, A(z) \to \forall z{\prec}y \oplus \omega^x\, A(z)).$$

We first show

$$\text{If } A(x) \text{ is progressive, then } A^+(x) \text{ is progressive,}$$

where "$B(x)$ is *progressive*" means $\forall x.\forall y{\prec}x\, B(y) \to B(x)$. So assume that $A(x)$ is progressive and

$$\forall y{\prec}x\, A^+(y). \tag{58}$$

We have to show $A^+(x)$. So assume further

$$\forall z{\prec}y\, A(z) \tag{59}$$

and $z \prec y \oplus \omega^x$. We have to show $A(z)$.

Case $x = 0$. Then $z \prec y \oplus \omega^0$. By (47) it suffices to derive $A(z)$ from $z \prec y$ as well as from $z = y$. If $z \prec y$, then $A(z)$ follows from (59), and if $z = y$, then $A(z)$ follows from (59) and the progressiveness of $A(x)$.

Case $\mathsf{S}x$. From $z \prec y \oplus \omega^{\mathsf{S}x}$ we obtain $z \prec y \oplus \omega^{e(x,y,z)} m(x,y,z)$ by (53) and $e(x,y,z) \prec \mathsf{S}x$ by (54). From (58) we obtain $A^+(e(x,y,z))$. By the definition of $A^+(x)$ we get

$$\forall u{\prec}y \oplus \omega^{e(x,y,z)} v\, A(u) \to \forall u{\prec}(y \oplus \omega^{e(x,y,z)} v) \oplus \omega^{e(x,y,z)}\, A(u)$$

and hence, using (49) and (52)

$$\forall u{\prec}y \oplus \omega^{e(x,y,z)} v\, A(u) \to \forall u{\prec}y \oplus \omega^{e(x,y,z)} \mathsf{S}(v)\, A(u).$$

Also from (59) and (51), (48) we obtain

$$\forall u{\prec}y \oplus \omega^{e(x,y,z)}0\, A(u).$$

Using an appropriate instance of the induction schema we can conclude

$$\forall u{\prec}y \oplus \omega^{e(x,y,z)}m(x,y,z)\, A(u)$$

and hence $A(z)$.

We now show, by induction on n, how for an arbitrary formula $A(x)$ we can obtain a derivation of

$$(\forall x.\forall y{\prec}x\, A(y) \to A(x)) \to \forall x{\prec}\omega_n\, A(x).$$

So assume the left hand side, i.e., assume that $A(x)$ is progressive.

Case 0. Then $x \prec \omega^0$ and hence $x \prec 0 \oplus \omega^0$ by (50). By (47) it suffices to derive $A(x)$ from $x \prec 0$ as well as from $x = 0$. Now $x \prec 0 \to A(x)$ holds by (46), and $A(0)$ then follows from the progressiveness of $A(x)$.

Case $n+1$. Since $A(x)$ is progressive, by what we have shown above $A^+(x)$ is also progressive. Applying the IH to $A^+(x)$ yields $\forall x{\prec}\omega_n\, A^+(x)$, and hence $A^+(\omega_n)$ by the progressiveness of $A^+(x)$. Now the definition of $A^+(x)$ (together with (46) and (50)) yields $\forall z{\prec}\omega^{\omega_n}\, A(z)$. □

Note that in the induction step of this proof we have derived transfinite induction up to ω_{n+1} for $A(x)$ from transfinite induction up to ω_n for a formula of level higher than the level of $A(x)$.

We now want to refine the preceeding theorem to a corresponding result for the subsystems $\mathbf{Z}_k$ of $\mathbf{Z}$.

THEOREM 2.2 (Provable Initial Cases of TI in $\mathbf{Z}_k$). *Let $1 \le l \le k$. Then in $\mathbf{Z}_k$ we can derive transfinite induction for any formula $A(x)$ of level $\le l$ up to $\omega_{k-l+2}[m]$ for arbitrary m, i.e.*

$$(\forall x.\forall y{\prec}x\, A(y) \to A(x)) \to \forall x{\prec}\omega_{k-l+2}[m]\, A(x),$$

where $\omega_1[m] := m$, $\omega_{i+1}[m] := \omega^{\omega_i[m]}$.

PROOF. Note first that if $A(x)$ is a formula of level $l \ge 1$, then the formula $A^+(x)$ constructed in the proof of the preceeding theorem has level $l+1$, and for the proof of

$$\text{If } A(x) \text{ is progressive, then } A^+(x) \text{ is progressive,}$$

we have used induction with an induction formula of level l.

Now let $A(x)$ be a fixed formula of level $\le l$, and assume that $A(x)$ is progressive. Define $A^0 := A$, $A^{i+1} := (A^i)^+$. Then $\mathsf{lev}(A^i) \le l+i$, and hence in $\mathbf{Z}_k$ we can derive that $A^1, A^2, \dots A^{k-l+1}$ are all progressive. Now from the progressiveness of $A^{k-l+1}(x)$ we obtain $A^{k-l+1}(0)$, $A^{k-l+1}(1)$, $A^{k-l+1}(2)$ and generally $A^{k-l+1}(m)$ for any m, i.e., $A^{k-l+1}(\omega_1[m])$. But since

$$A^{k-l+1}(x) = (A^{k-l})^+(x) = \forall y(\forall z{\prec}y\, A^{k-l}(z) \to \forall z{\prec}y \oplus \omega^x\, A^{k-l}(z))$$

we first get (with $y = 0$) $\forall z{\prec}\omega_2[m]\, A^{k-l}(z)$ and then $A^{k-l}(\omega_2[m])$ by the progressiveness of A^{k-l}. Repeating this argument we finally obtain

$$\forall z{\prec}\omega_{k-l+2}[m]\, A^0(z).$$

This concludes the proof. □

Our next aim is to prove that these bounds are sharp. More precisely, we will show that in $\mathbf{Z}$ (no matter how many true Π_1-formulas we have added as axioms) one cannot derive "purely schematic" transfinite induction up to ε_0, i.e., one cannot derive the formula

$$(\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x Px$$

with a relation symbol P, and that in $\mathbf{Z}_k$ one cannot derive transfinite induction up to ω_{k+1}, i.e., the formula

$$(\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x{\prec}\omega_{k+1}\, Px.$$

This will follow from the method of normalization applied to arithmetical systems, which we have to develop first.

3. Normalization with the Omega Rule

We will show in Theorem 4.7 that a normalization theorem does not hold for arithmetical systems $\mathbf{Z}$, in the sense that for any formula A derivable in $\mathbf{Z}$ there is a derivation of the same formula A in $\mathbf{Z}$ which only uses formulas of a level bounded by the level of A. The reason for this failure is the presence of induction axioms, which can be of arbitrary level.

Here we remove that obstacle against normalization in a somewhat drastic way: we leave the realm of proofs as finite combinatory objects and replace the induction axiom by a rule with infinitely many premises, the so-called ω-rule (suggested by Hilbert and studied by Lorenzen, Novikov and Schütte), which allows us to conclude $\forall x A(x)$ from $A(0), A(1), A(2), \ldots$, i.e.

$$\frac{\begin{array}{cccccc} d_0 & d_1 & & d_i & \\ A(0) & A(1) & \ldots & A(i) & \ldots \end{array}}{\forall x A(x)}\,\omega$$

So derivations can be viewed as labelled infinite (countably branching) trees. As in the finitary case a label consists of the derived formula and the name of the rule applied. Since we define derivations inductively, any such derivation tree must be well-founded, i.e., must not contain an infinite descending path.

Clearly this ω-rule can also be used to replace the rule $\forall^+ x$. As a consequence we do not need to consider free individual variables.

It is plain that every derivation in an arithmetical system $\mathbf{Z}$ can be translated into an infinitary derivation with the ω-rule; this will be carried out in Lemma 3.3 below. The resulting infinitary derivation has a noteworthy property: in any application of the ω-rule the cutranks of the infinitely many immediate subderivations d_n are bounded, and also their sets of free assumption variables are bounded by a finite set. Here the cutrank of a derivation is as usual the least number $\geq$ the level of any subderivation obtained by $\to^+$ as the main premise of $\to^-$ or by the ω-rule as the main premise of $\forall^-$, where the *level of a derivation* is the level of its type as a term, i.e., of the formula it derives. Clearly a derivation is called normal iff its cutrank is zero, and we will prove below that any (possibly infinite) derivation of finite cutrank can be transformed into a derivation of cutrank zero. The resulting normal derivation will continue to be infinite, so the result may seem useless at first sight. However, we will be able to bound the depth of the resulting derivation in an informative way, and this will enable

us in Section 4 to obtain the desired results on unprovable initial cases of transfinite induction. Let us now carry out this programme.

N.B. The standard definition of cutrank in predicate logic measures the depth of formulas; here one uses the level.

3.1. Infinitary Derivations. The systems $\mathbf{Z}^\infty$ of ω-arithmetic are defined as follows. $\mathbf{Z}^\infty$ has the same language and – apart from the induction axioms – the same axioms as $\mathbf{Z}$. Derivations in $\mathbf{Z}^\infty$ are infinite objects. It is useful to employ a term notation for these, and we temporarily use d, e, f to denote such (infinitary) derivation terms. For the term corresponding to the deduction obtained by applying the ω-rule to d_i, $i \in \mathbb{N}$ we write $\langle d_i\rangle_{i<\omega}$. However, for our purposes here it suffices to only consider derivations whose depth is bounded below ε_0.

We define the notion "d is a *derivation of depth* $\leq \alpha$" (written $|d| \leq \alpha$) inductively as follows (i ranges over numerals).

(A) Any assumption variable u^A with A a closed formula and any axiom Ax^A is a derivation of depth $\leq \alpha$, for any α.

($\to^+$) If d^B is a derivation of depth $\leq \alpha_0 < \alpha$, then $(\lambda u^A.d^B)^{A\to B}$ is a derivation of depth $\leq \alpha$.

($\to^-$) If $d^{A\to B}$ and e^A are derivations of depths $\leq \alpha_i < \alpha$ (i=1,2), then $(d^{A\to B}e^A)^B$ is a derivation of depth $\leq \alpha$.

(ω) For all $A(x)$, if $d_i^{A(i)}$ are derivations of depths $\leq \alpha_i < \alpha$ ($i < \omega$), then $(\langle d_i^{A(i)}\rangle_{i<\omega})^{\forall x A}$ is a derivation of depth $\leq \alpha$.

($\forall^-$) For all $A(x)$, if $d^{\forall x A(x)}$ is a derivation of depth $\leq \alpha_0 < \alpha$, then, for all i, $(d^{\forall x A(x)}i)^{A(i)}$ is a derivation of depth $\leq \alpha$.

We will use $|d|$ to denote the least α such that $|d| \leq \alpha$.

Note that in ($\forall^-$) it suffices to use numerals as minor premises. The reason is that we only need to consider closed terms, and any such term is in our setup identified with a numeral.

The *cutrank* $\mathsf{cr}(d)$ of a derivation d is defined by

$$\begin{aligned}
\mathsf{cr}(u^A) &:= \mathsf{cr}(\mathsf{Ax}^A) := 0,\\
\mathsf{cr}(\lambda u d) &:= \mathsf{cr}(d),\\
\mathsf{cr}(d^{A\to B}e^A) &:= \begin{cases} \max(\mathsf{lev}(A\to B), \mathsf{cr}(d), \mathsf{cr}(e)), & \text{if } d = \lambda u d',\\ \max(\mathsf{cr}(d), \mathsf{cr}(e)), & \text{otherwise,}\end{cases}\\
\mathsf{cr}(\langle d_i\rangle_{i<\omega}) &:= \sup_{i<\omega} \mathsf{cr}(d_i),\\
\mathsf{cr}(d^{\forall x A(x)}j) &:= \begin{cases} \max(\mathsf{lev}(\forall x A(x)), \mathsf{cr}(d)), & \text{if } d = \langle d_i\rangle_{i<\omega},\\ \mathsf{cr}(d), & \text{otherwise.}\end{cases}
\end{aligned}$$

Clearly $\mathsf{cr}(d) \in \mathbb{N} \cup \{\omega\}$ for all d. For our purposes it will suffice to consider only derivations with finite cutranks (i.e., with $\mathsf{cr}(d) \in \mathbb{N}$) and with finitely many free assumption variables.

LEMMA 3.1. *If d is a derivation of depth $\leq \alpha$, with free assumption variables among $u, \vec{u}$ and of cutrank $\mathsf{cr}(d) = k$, and e is a derivation of depth $\leq \beta$, with free assumption variables among $\vec{u}$ and of cutrank $\mathsf{cr}(e) = l$,*

then $d[u := e]$ *is a derivation with free assumption variables among* $\vec{u}$*, of depth* $|d[u := e]| \leq \beta + \alpha$ *and of cutrank* $\mathsf{cr}(d[u := e]) \leq \max(\mathsf{lev}(e), k, l)$.

PROOF. Straightforward induction on the depth of d. □

Using this lemma we can now embed our systems $\mathbf{Z}_k$ (i.e., arithmetic with induction restricted to formulas of level $\leq k$) and hence $\mathbf{Z}$ into $\mathbf{Z}^\infty$. In this embedding we refer to the number $n_I(d)$ of nested applications of the induction schema within a $\mathbf{Z}_k$-derivation d.

The *nesting* of applications of induction in d, $n_I(d)$, is defined by induction on d, as follows.

$$\begin{aligned}
n_I(u) &:= n_I(\mathsf{Ax}) := 0,\\
n_I(\mathsf{Ind}) &:= 1,\\
n_I(\mathsf{Ind}\,\vec{t}de) &:= \max(n_I(d), n_I(e) + 1),\\
n_I(de) &:= \max(n_I(d), n_I(e)), \quad \text{if } d \text{ is not of the form } \mathsf{Ind}\,\vec{t}d_0,\\
n_I(\lambda u d) &:= n_I(\lambda x d) := n_I(dt) := n_I(d).
\end{aligned}$$

3.2. Long Normal Form. For the next lemma we need the notion of the *long normal form* of a derivation. In Subsection 3.7 of Chapter 1 we have studied the form of normal derivations in minimal logic. We considered the notion of a *track* and observed, that in every track all elimination rules precede all introduction rules, and that in a uniquely determined *minimal node* we encounter a *minimal formula*, that is a subformula of any formula in the elimination part as well as in the introduction part of the track. In the notion of a long normal form we additionally require that every minimal formula is atomic.

For simplicity we restrict ourselves to the $\to$-fragment of minimal propositional logic; however, our considerations are valid for the full language as well.

For terms of the typed λ-calculus we define the *η-expansion of a variable* by

$$\eta_V(x^{\vec{\tau}\to\iota}) := \lambda\vec{z}^{\vec{\tau}}.x\eta_V(\vec{z}),$$

so by induction on the type of the variable. The *η-expansion of a term* can then be defined by induction on terms:

$$\eta\big(\lambda\vec{y}.(x\vec{M})^{\vec{\tau}\to\iota}\big) := \lambda\vec{y},\vec{z}^{\vec{\tau}}.x\eta(\vec{M})\eta_V(\vec{z}).$$

Note that we always have $\eta(x) = \eta_V(x)$. – Hence clearly:

LEMMA 3.2. *Every term can be transformed into long normal form, by first normalizing and then η-expanding it.*

3.3. Embedding of $\mathbf{Z}_k$.

LEMMA 3.3. *Let a $\mathbf{Z}_k$-derivation in long normal form be given with $\leq m$ nested applications of the induction schema, i.e., of*

$$A(0) \to (\forall x.A(x) \to A(\mathsf{S}x)) \to \forall x A(x),$$

all with $\mathsf{lev}(A) \leq k$. *We consider subderivations d^B not of the form $\mathsf{Ind}\,\vec{t}$ or $\mathsf{Ind}\,\vec{t}d_0$. For every such subderivation and closed substitution instance $B\sigma$ of B we construct $(d^\infty_\sigma)^{B\sigma}$ in $\mathbf{Z}^\infty$ with free assumption variables $u^{C\sigma}$ for u^C*

free assumption of d, such that $|d^\infty_\sigma| < \omega^{m+1}$ and $\mathsf{cr}(d^\infty_\sigma) \le k$, and moreover such that d is obtained by $\to^+$ iff d^∞_σ is, and d is obtained by $\forall^+$ or of the form $\mathsf{Ind}\,\vec{t}d_0e$ iff d^∞_σ is obtained by the ω-rule.

PROOF. By recursion on such subderivations d.

Case u^C or Ax. Take $u^{C\sigma}$ or Ax.

Case $\mathsf{Ind}\,\vec{t}de'$. Since the deduction is in long normal form, $e' = \lambda xv.e$. By IH we have d^∞_σ and e^∞_σ. (Note that neither d nor e can have one of the forbidden forms $\mathsf{Ind}\,\vec{t}$ and $\mathsf{Ind}\,\vec{t}d_0$, since both are in long normal form). Write $e^\infty_\sigma(t,f)$ for $e^\infty_\sigma[x,v := t,f]$, and let

$$(\mathsf{Ind}\,\vec{t}d(\lambda xv.e))^\infty_\sigma := \langle d^\infty_\sigma, e^\infty_\sigma(0,d^\infty_\sigma), e^\infty_\sigma(1,e^\infty_\sigma(0,d^\infty_\sigma)),\dots\rangle.$$

By IH $|e^\infty_\sigma| \le \omega^{m-1}\cdot p$ and $|d^\infty_\sigma| \le \omega^m\cdot q$ for some $p,q<\omega$. By Lemma 3.1 we obtain

$$|e^\infty_\sigma(0,d^\infty_\sigma)| \le \omega^m\cdot q + \omega^{m-1}\cdot p,$$
$$|e^\infty_\sigma(1,e^\infty_\sigma(0,d^\infty_\sigma))| \le \omega^m\cdot q + \omega^{m-1}\cdot 2p$$

and so on, and hence

$$|(\mathsf{Ind}\,d(\lambda xv.e))^\infty_\sigma| \le \omega^m\cdot(q+1).$$

Concerning the cutrank we have by IH $\mathsf{cr}(d^\infty_\sigma), \mathsf{cr}(e^\infty_\sigma) \le k$. Therefore

$$\mathsf{cr}(e^\infty_\sigma(0,d^\infty_\sigma)) \le \max(\mathsf{lev}(A(0)), \mathsf{cr}(d^\infty_\sigma), \mathsf{cr}(e^\infty_\sigma)) \le k,$$
$$\mathsf{cr}(e^\infty_\sigma(1,e^\infty_\sigma(0,d^\infty_\sigma))) \le \max(\mathsf{lev}(A(1)), k, \mathsf{cr}(e^\infty_\sigma)) = k,$$

and so on, and hence

$$\mathsf{cr}((\mathsf{Ind}\,d(\lambda xv.e))^\infty_\sigma) \le k.$$

Case $\lambda u^C.d^B$. By IH, we have $(d^\infty_\sigma)^{B\sigma}$ with possibly free assumptions $u^{C\sigma}$. Take $(\lambda u.d)^\infty_\sigma := \lambda u^{C\sigma}.d^\infty_\sigma$.

Case de, with d not of the form $\mathsf{Ind}\,\vec{t}$ or $\mathsf{Ind}\,\vec{t}d_0$. By IH we have d^∞_σ and e^∞_σ. Since de is subderivation of a normal derivation we know that d and hence also d^∞_σ is not obtained by $\to^+$. Therefore $(de)^\infty_\sigma := d^\infty_\sigma e^\infty_\sigma$ is normal and $\mathsf{cr}(d^\infty_\sigma e^\infty_\sigma) = \max(\mathsf{cr}(d^\infty_\sigma), \mathsf{cr}(e^\infty_\sigma)) \le k$. Also we clearly have $|d^\infty_\sigma e^\infty_\sigma| < \omega^{m+1}$.

Case $(\lambda x.d)^{\forall xB(x)}$. By IH for every i and substitution instance $B(i)\sigma$ we have $d^\infty_{\sigma,i}$. Take $(\lambda x.d)^\infty_\sigma := \langle d^\infty_{\sigma,i}\rangle_{i<\omega}$.

Case $(dt)^{B[x:=t]}$. By IH, we have $(d^\infty_\sigma)^{(\forall xB)\sigma}$. Let j be the numeral with the same value as $t\sigma$. If $d^\infty_\sigma = \langle d_i\rangle_{i<\omega}$ (which can only be the case if $d = \mathsf{Ind}\,\vec{t}d_0e_0$, for dt is a subderivation of a normal derivation), take $(dt)^\infty_\sigma := d_j$. Otherwise take $(dt)^\infty_\sigma := d^\infty_\sigma j$ □

3.4. Normalization for $\mathbf{Z}^\infty$. A derivation is called *convertible* or a *redex* if it is of the form $(\lambda u.d)e$ or else $\langle d_i\rangle_{i<\omega}j$, which can be converted into $d[u := e]$ or d_j, respectively. A derivation is called *normal* if it does not contain a convertible subderivation. Note that a derivation is normal iff it is of cutrank 0.

Call a derivation a *simple application* if it is of the form $d_0d_1\dots d_m$ with d_0 an assumption variable or an axiom.

We want to define an operation which by repeated conversions transforms a given derivation into a normal one with the same end formula and

no additional free assumption variables. The usual methods to achieve such a task have to be adapted properly in order to deal with the new situation of infinitary derivations. Here we give a particularly simple argument due to Tait [**26**].

LEMMA 3.4. *For any derivation d^A of depth $\leq \alpha$ and cutrank $k+1$ we can find a derivation $(d^k)^A$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha$ and cutrank $\leq k$.*

PROOF. By induction on α. The only case which requires some argument is when the derivation is of the form de with $|d| \leq \alpha_1 < \alpha$ and $|e| \leq \alpha_2 < \alpha$, but is not a simple application. We first consider the subcase where $d^k = \lambda u.d_1(u)$ and $\mathsf{lev}(d) = k+1$. Then $\mathsf{lev}(e) \leq k$ by the definition of level (recall that the level of a derivation was defined to be the level of the formula it derives), and hence $d_1[u := e^k]$ has cutrank $\leq k$ by Lemma 3.1. Furthermore, also by Lemma 3.1, $d_1[u := e^k]$ has depth $\leq 2^{\alpha_2} + 2^{\alpha_1} \leq 2^{\max(\alpha_2,\alpha_1)+1} \leq 2^\alpha$. Hence we can take $(de)^k$ to be $d_1[u := e^k]$.

In the subcase where $d^k = \langle d_i \rangle_{i<\omega}$, $\mathsf{lev}(d) = k+1$ and $e^k = j$ we can take $(de)^k$ to be d_j, since clearly d_j has cutrank $\leq k$ and depth $\leq 2^\alpha$. If we are not in the above subcases, we can simply take $(de)^k$ to be $d^k e^k$. This derivation clearly has depth $\leq 2^\alpha$. Also it has cutrank $\leq k$, which can be seen as follows. If $\mathsf{lev}(d) \leq k+1$ we are done. But $\mathsf{lev}(d) \geq k+2$ is impossible, since we have assumed that de is not a simple application. In order to see this, note that if de is not a simple application, it must be of the form $d_0 d_1 \dots d_n e$ with d_0 not an assumption variable or axiom and d_0 not itself of the form $d'd''$; then d_0 must end with an introduction $\to^+$ or ω, hence there is a cut of a degree exceeding $k+1$, which is excluded by assumption. □

As an immediate consequence we obtain:

THEOREM 3.5 (Normalization for $\mathbf{Z}^\infty$). *For any derivation d^A of depth $\leq \alpha$ and cutrank $\leq k$ we can find a normal derivation $(d^*)^A$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha_k$, where $2^\alpha_0 := \alpha$, $2^\alpha_{m+1} := 2^{2^\alpha_m}$.*

As in Section 3.7 of Chapter 1 we can now analyze the structure of normal derivations in $\mathbf{Z}^\infty$. In particular we obtain:

THEOREM 3.6 (Subformula Property for $\mathbf{Z}^\infty$). *Let d be a normal deduction in $\mathbf{Z}^\infty$ for $\Gamma \vdash A$. Then each formula in d is a subformula of a formula in $\Gamma \cup \{A\}$.*

PROOF. We prove this for tracks of order n, by induction on n. □

4. Unprovable Initial Cases of Transfinite Induction

We now apply the technique of normalization for arithmetic with the ω-rule to obtain a proof that transfinite induction up to ε_0 is underivable in $\mathbf{Z}$, i.e., a proof of

$$\mathbf{Z} \nvdash (\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x Px$$

with a relation symbol P, and that transfinite induction up to ω_{k+1} is underivable in $\mathbf{Z}_k$, i.e., a proof of

$$\mathbf{Z}_k \not\vdash (\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x{\prec}\omega_{k+1}\, Px.$$

It clearly suffices to prove this for arithmetical systems based on classical logic. Hence we may assume that we have used only the classical versions (55), (56) and (57) of the axioms from Subsection 2.1.

Our proof is based on an idea of Schütte, which consists in adding a so-called *progression rule* to the infinitary systems. This rule allows us to conclude Pj (where j is any numeral) from all Pi for $i \prec j$.

4.1. Progression Rule. More precisely, we define the notion of a derivation in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ of depth $\leq \alpha$ by the inductive clauses above and the additional clause $\mathsf{Prog}(P)$:

(Prog) If for all $i \prec j$ we have derivations d_i^{Pi} of depths $\leq \alpha_i < \alpha$, then $\langle d_i^{Pi}\rangle_{i\prec j}^{Pj}$ is a derivation of depth $\leq \alpha$.

We also define $\mathsf{cr}(\langle d_i\rangle_{i\prec j}) := \sup_{i\prec j} \mathsf{cr}(d_i)$.

Since this progression rule only deals with derivations of atomic formulas, it does not affect the cutranks of derivations. Hence the proof of normalization for $\mathbf{Z}^\infty$ carries over unchanged to $\mathbf{Z}^\infty + \mathsf{Prog}(P)$. In particular we have

LEMMA 4.1. *For any derivation d^A in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ of depth $\leq \alpha$ and cutrank $\leq k+1$ we can find a derivation $(d^k)^A$ in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ with free assumption variables contained in those of d, which has depth $\leq 2^\alpha$ and cutrank $\leq k$.*

We now show that from the progression rule for P we can easily derive the progressiveness of P.

LEMMA 4.2. *We have a normal derivation of $\forall x.\forall y{\prec}x\, Py \to Px$ in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ with depth 5.*

PROOF.

$$\dfrac{\dfrac{\dots \quad \dfrac{\dfrac{\forall y{\prec}j\, Py}{i\prec j \to Pi}\forall^- \qquad i\prec j}{Pi}\to^- \quad \dots \quad (\text{all } i\prec j)}{Pj}\mathit{Prog} \qquad \dots \quad (\text{all } j)}{\forall x.\forall y{\prec}x\, Py \to Px}\,\omega$$
$$\forall y{\prec}j\, Py \to Pj\ \to^+$$

□

4.2. Quasi-Normal Derivations. The crucial observation now is that a normal derivation of $P^\ulcorner\beta^\urcorner$ must essentially have a depth of at least β. However, to obtain the right estimates for the subsystems $\mathbf{Z}_k$ we cannot apply Lemma 4.1 down to cutrank 0 (i.e., to normal form) but must stop at cutrank 1. Such derivations, i.e., those of cutrank ≤ 1, will be called *quasi-normal*; they can also be analyzed easily.

We begin by showing that a quasi-normal derivation of a quantifier-free formula can always be transformed without increasing its cutrank or its depth into a quasi-normal derivation of the same formula which

(1) does not use the ω-rule, *and*
(2) contains $\forall^-$ only in the initial part of a track starting with an axiom.

Recall that our axioms are of the form $\forall \vec{x} A$ with A quantifier-free.

The *quasi-subformulas* of a formula A are defined by the clauses

(a) A, B are quasi-subformulas of $A \to B$;
(b) $A(i)$ is a quasi-subformula of $\forall x A(x)$, for all numerals i;
(c) If A is a quasi-subformula of B, and C is an atomic formula, then $C \to A$ and $\forall x A$ are quasi-subformulas of B;
(d) "... is quasi-subformula of ..." is a reflexive and transitive relation.

For example, $Q \to \forall x.P \to A$, P, Q atomic, is a quasi-subformula of $A \to B$.

We now transfer the subformula property for normal derivations (Theorem 3.6) to a quasi-subformula property for quasi-normal derivations.

THEOREM 4.3 (Quasi-Subformula Property). *Let d be a quasi-normal deduction in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ for $\Gamma \vdash A$. Then each formula in d is a quasi-subformula of a formula in $\Gamma \cup \{A\}$.*

PROOF. We prove this for tracks of order n, by induction on n. □

COROLLARY 4.4. *Let d be a quasi-normal deduction in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ of a formula $\forall \vec{x} A$ with A quantifier-free from quantifier-free assumptions. Then any track in d of positive order ends with a quantifier-free formula.*

PROOF. If not, then the major premise of the $\to^-$ whose minor premise is the offending end formula of the track, would contain a quantifier to the left of $\to$. This contradicts Theorem 4.3. □

4.3. Elimination of the Omega Rule. Our next aim is to eliminate the ω-rule. For this we need the notion of an *instance* of a formula, defined by the following clauses.

(a) If B' is an instance of B and A is quantifier-free, then $A \to B'$ is an instance of $A \to B$;
(b) $A(i)$ is an instance of $\forall x A(x)$, for all numerals i;
(c) The relation "... is an instance of ..." is reflexive and transitive.

LEMMA 4.5. *Let d be a quasi-normal deduction in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ of a formula A without $\forall$ to the left of $\to$ from quantifier-free assumptions. Then for any quantifier-free instance A' of A we can find a quasi-normal derivation d' of A' from the same assumptions such that*

(a) *d' does not use the ω-rule,*
(b) *d' contains $\forall^-$ only in the initial elimination part of a track starting with an axiom, and*
(c) *$|d'| \leq |d|$.*

PROOF. By induction on the depth of d. We distinguish cases according to the last rule in d.

Case $\to^-$.

$$\frac{A \to B \qquad A}{B} \to^-$$

By the quasi-subformula property A must be quantifier-free. Let B' be a quantifier-free instance of B. Then by definition $A \to B'$ is a quantifier-free instance of $A \to B$. The claim now follows from the IH.

Case $\to^+$.

$$\frac{B}{A \to B}\,{\to^+}$$

Any instance of $A \to B$ has the form $A \to B'$ with B' an instance of B. Hence the claim follows from the IH.

Case $\forall^-$.

$$\frac{\forall x A(x) \qquad i}{A(i)}\,\forall^-$$

Then any quantifier-free instance of $A(i)$ is also a quantifier-free instance of $\forall x A(x)$, and hence the claim follows from the IH.

Case ω.

$$\frac{\ldots \quad A(i) \quad \ldots \quad (\text{all } i < \omega)}{\forall x A(x)}\,\omega$$

Any quantifier-free instance of $\forall x A(x)$ has the form $A(i)'$ with $A(i)'$ a quantifier-free instance of $A(i)$. Hence the claim again follows from the IH. □

A derivation d in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ is called a $P\vec{\alpha}, \neg P\vec{\beta}$-*refutation* if $\vec{\alpha}$ and $\vec{\beta}$ are disjoint and d derives a formula $\vec{A} \to B := A_1 \to \cdots \to A_k \to B$ with $\vec{A}$ and the free assumptions in d among $P^\ulcorner \alpha_1 \urcorner, \ldots, P^\ulcorner \alpha_m \urcorner, \neg P^\ulcorner \beta_1 \urcorner, \ldots, \neg P^\ulcorner \beta_n \urcorner$ or true quantifier-free formulas without P, and B a false quantifier-free formula without P or else among $P^\ulcorner \beta_1 \urcorner, \ldots, P^\ulcorner \beta_n \urcorner$.

(So, classically, a $P\vec{\alpha}, \neg P\vec{\beta}$-refutation shows $\bigwedge_i P^\ulcorner \alpha_i \urcorner \to \bigvee_j P^\ulcorner \beta_j \urcorner$.)

Lemma 4.6. *Let d be a quasi-normal $P\vec{\alpha}, \neg P\vec{\beta}$-refutation. Then*

$$\min(\vec{\beta}) \le |d| + \mathsf{lh}(\vec{\alpha}'),$$

where $\vec{\alpha}'$ is the sublist of $\vec{\alpha}$ consisting of all $\alpha_i < \min(\vec{\beta})$, and $\mathsf{lh}(\vec{\alpha}')$ denotes the length of the list $\vec{\alpha}'$.

Proof. By induction on $|d|$. By the Lemma above we may assume that d does not contain the ω-rule, and contains $\forall^-$ only in a context where leading universal quantifiers of an axiom are removed. We distinguish cases according to the last rule in d.

Case $\to^+$. By our definition of refutations the claim follows immediately from the IH.

Case $\to^-$. Then $d = f^{C \to (\vec{A} \to B)} e^C$. If C is a true quantifier-free formula without P or of the form $P^\ulcorner \gamma \urcorner$ with $\gamma < \min(\vec{\beta})$, the claim follows from the IH for f:

$$\min(\vec{\beta}) \le |f| + \mathsf{lh}(\vec{\alpha}') + 1 \le |d| + \mathsf{lh}(\vec{\alpha}').$$

If C is a false quantifier-free formula without P or of the form $P^\ulcorner \gamma \urcorner$ with $\min(\vec{\beta}) \le \gamma$, the claim follows from the IH for e:

$$\min(\vec{\beta}) \le |e| + \mathsf{lh}(\vec{\alpha}') + 1 \le |d| + \mathsf{lh}(\vec{\alpha}').$$

It remains to consider the case when C is a quantifier-free implication involving P. Then $\mathsf{lev}(C) \ge 1$, hence $\mathsf{lev}(C \to (\vec{A} \to B)) \ge 2$ and therefore

(since $\mathsf{cr}(d) \leq 1$) f must be a simple application starting with an axiom. Now our only axioms involving P are $\mathsf{Eq}_P\colon \forall x, y.x = y \to Px \to Py$ and Stab_P *colon*$\forall x.\neg\neg Px \to Px$, and of these only Stab_P has the right form. Hence $f = \mathsf{Stab}_P{}^{\ulcorner}\gamma^{\urcorner}$ and therefore $e\colon \neg\neg P^{\ulcorner}\gamma^{\urcorner}$. Now from $\mathsf{lev}(\neg\neg P^{\ulcorner}\gamma^{\urcorner}) = 2$, the assumption $\mathsf{cr}(e) \leq 1$ and again the form of our axioms involving P, it follows that e must end with $\to^+$, i.e., $e = \lambda u^{\neg P^{\ulcorner}\gamma^{\urcorner}}.e_0^{\bot}$. So we have

$$\dfrac{\dfrac{f}{\neg\neg P^{\ulcorner}\gamma^{\urcorner} \to P^{\ulcorner}\gamma^{\urcorner}} \qquad \dfrac{\begin{array}{c}[u\colon \neg P^{\ulcorner}\gamma^{\urcorner}] \\ e_0 \\ \bot\end{array}}{\neg\neg P^{\ulcorner}\gamma^{\urcorner}}}{P^{\ulcorner}\gamma^{\urcorner}}$$

The claim now follows from the IH for e_0.

Case $\forall^-$. By assumption we then are in the initial part of a track starting with an axiom. Since d is a $P\vec{\alpha}, \neg P\vec{\beta}$-refutation, that axiom must contain P. It cannot be the equality axiom $\mathsf{Eq}_P\colon \forall x, y.x = y \to Px \to Py$, since $^{\ulcorner}\gamma^{\urcorner} = {}^{\ulcorner}\delta^{\urcorner} \to P^{\ulcorner}\gamma^{\urcorner} \to P^{\ulcorner}\delta^{\urcorner}$ can never be (whether $\gamma = \delta$ or $\gamma \neq \delta$) the end formula of a $P\vec{\alpha}, \neg P\vec{\beta}$-refutation. For the same reason it can not be the stability axiom $\mathsf{Stab}_P\colon \forall x.\neg\neg Px \to Px$). Hence the case $\forall^-$ cannot occur.

Case $\mathrm{Prog}(P)$. Then $d = \langle d_\delta^{P^{\ulcorner}\delta^{\urcorner}}\rangle_{\delta<\gamma}^{P^{\ulcorner}\gamma^{\urcorner}}$. By assumption on d, γ is in $\vec{\beta}$. We may assume $\gamma = \beta_i := \min(\vec{\beta})$, for otherwise the premise deduction $d_{\beta_i}\colon P^{\ulcorner}\beta_i{}^{\urcorner}$ would be a quasi-normal $P\vec{\alpha}, \neg P\vec{\beta}$-refutation, to which we could apply the IH.

If there are no $\alpha_j < \gamma$, then the argument is simple: every d_δ is a $P\vec{\alpha}, \neg P\vec{\beta}, \neg P\delta$-refutation, so by IH, since also no $\alpha_j < \delta$,

$$\min(\vec{\beta}, \delta) = \delta \leq \mathsf{dp}(d_\delta),$$

hence $\gamma = \min(\vec{\beta}) \leq |d|$.

To deal with the situation that some α_j are less than γ, we observe that there can be at most finitely many α_j immediately preceeding γ; so let ε be the least ordinal such that

$$\forall\delta.\varepsilon \leq \delta < \gamma \to \delta \in \vec{\alpha}.$$

Then $\varepsilon, \varepsilon+1, \ldots, \varepsilon+k-1 \in \vec{\alpha}$, $\varepsilon + k = \gamma$. We may assume that ε is either a successor or a limit. If $\varepsilon = \varepsilon' + 1$, it follows by the IH that since $d_{\varepsilon'}$ is a $P\vec{\alpha}, \neg P\vec{\beta}, \neg P(\varepsilon-1)$-refutation,

$$\varepsilon - 1 \leq \mathsf{dp}(d_{\varepsilon-1}) + \mathsf{lh}(\vec{\alpha}') - k,$$

where $\vec{\alpha}'$ is the sequence of $\alpha_j < \gamma$. Hence $\varepsilon \leq |d| + \mathsf{lh}(\vec{\alpha}') - k$, and so

$$\gamma \leq |d| + \mathsf{lh}(\vec{\alpha}').$$

If ε is a limit, there is a sequence $\langle\delta_{f(n)}\rangle_n$ with limit ε, and with all $\alpha_j < \varepsilon$ below $\delta_{f(0)}$, and so by IH

$$\delta_{f(n)} \leq \mathsf{dp}(d_{f(n)}) + \mathsf{lh}(\vec{\alpha}') - k,$$

and hence $\varepsilon \leq |d_\varepsilon| + \mathsf{lh}(\vec{\alpha}') - k$, so $\gamma \leq |d| + \mathsf{lh}(\vec{\alpha}')$. □

4.4. Underivability of Transfinite Induction.

THEOREM. *Transfinite induction up to ε_0 is underivable in $\mathbf{Z}$, i.e.*

$$\mathbf{Z} \nvdash (\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x Px$$

with a relation symbol P, and for $k \geq 3$ transfinite induction up to ω_{k+1} is underivable in $\mathbf{Z}_k$, i.e.,

$$\mathbf{Z}_k \nvdash (\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x{\prec}\omega_{k+1}\, Px.$$

PROOF. We restrict ourselves to the second part. So assume that transfinite induction up to ω_{k+1} is derivable in $\mathbf{Z}_k$. Then by the embedding of $\mathbf{Z}_k$ into $\mathbf{Z}^\infty$ and the normal derivability of the progressiveness of P in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ with finite depth we can conclude that $\forall x{\prec}\omega_{k+1}\, Px$ is derivable in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ with depth $< \omega^{m+1}$ and cutrank $\leq k$. (Note that here we need $k \geq 3$, since the formula expressing transfinite induction up to ω_{k+1} has level 3). Now $k-1$ applications of Lemma 4.1 yield a derivation of the same formula $\forall x{\prec}\omega_{k+1}\, Px$ in $\mathbf{Z}^\infty + \mathsf{Prog}(P)$ with depth $\gamma < 2_{k-1}^{\omega^{m+1}} < \omega_{k+1}$ and cutrank ≤ 1.

Hence there is also a quasi-normal derivation of $P^\ulcorner\gamma + 3^\urcorner$ in $\mathbf{Z}^\infty$ + Prog(P) with depth $\gamma + 2$ and cutrank ≤ 1, of the form

$$\dfrac{\dfrac{\begin{matrix}d\\ \forall x{\prec}\omega_{k+1} Px\end{matrix}}{{}^\ulcorner\gamma + 3^\urcorner \prec \omega_{k+1} \to P^\ulcorner\gamma + 3^\urcorner} \qquad \begin{matrix}d'\\ {}^\ulcorner\gamma + 3^\urcorner \prec \omega_{k+1}\end{matrix}}{P^\ulcorner\gamma + 3^\urcorner}$$

where d' is a deduction of finite depth (it may even be an axiom, depending on the precise choice of axioms for $\mathbf{Z}$); this contradicts the lemma just proved. □

4.5. Normalization for Arithmetic is Impossible. The normalization theorem for first-order logic applied to one of our arithmetical systems $\mathbf{Z}$ is not particularly useful since we may have used in our derivation induction axioms of arbitrary complexity. Hence it is tempting to first eliminate the induction schema in favour of an induction rule allowing us to conclude $\forall x A(x)$ from a derivation of $A(0)$ and a derivation of $A(\mathsf{S}x)$ with an additional assumption $A(x)$ to be cancelled at this point (note that this rule is equivalent to the induction schema), and then to try to normalize the resulting derivation in the new system $\mathbf{Z}$ with the induction rule. We will apply Gentzen's Theorems on Underivability and Derivability of Transfinite Induction to show that even a very weak form of the normalization theorem cannot hold in $\mathbf{Z}$ with the induction rule.

THEOREM. *The following weak form of a normalization theorem for $\mathbf{Z}$ with the induction rule is false: "For any derivation d^B with free assumption variables among $\vec{u}^{\vec{A}}$ for formulas $\vec{A}, B$ of level $\leq l$ there is a derivation $(d^*)^B$, with free assumption variables contained in those of d, which contains only formulas of level $\leq k$, where k depends on l only."*

PROOF. Assume that such a normalization theorem holds. Consider the formula

$$(\forall x.\forall y{\prec}x\, Py \to Px) \to \forall x{\prec}\omega_{n+1}\, Px$$

expressing transfinite induction up to ω_{n+1}, which is of level 3. By Gentzen's Theorems on Derivability of Transfinite Induction it is derivable in $\mathbf{Z}$. Now from our assumption it follows that there exists a derivation of this formula containing only formulas of level $\leq k$, for some k independent of n. Hence $\mathbf{Z}_k$ derives transfinite induction up to ω_{n+1} for any n. But this clearly contradicts theorem above (Underivability of Transfinite Induction). $\square$

Bibliography

1. E.W. Beth, *Semantic construction of intuitionistic logic*, Medelingen de KNAW N.S. **19** (1956), no. 11.
2. ———, *The foundations of mathematics*, North–Holland, Amsterdam, 1959.
3. Egon Börger, Erich Grädel, and Yuri Gurevich, *The classical decision problem*, Perspectives in Mathematical Logic, Springer Verlag, Berlin, Heidelberg, New York, 1997.
4. Wilfried Buchholz, *A new system of proof–theoretic ordinal functions*, Annals of Pure and Applied Logic **32** (1986), no. 3, 195–207.
5. Georg Cantor, *Beträge zur Begründung der transfiniten Mengenlehre*, Mathematische Annalen **49** (1897).
6. C.C. Chang and H.J. Keisler, *Model theory*, 3rd ed., Studies in Logic, vol. 73, North–Holland, Amsterdam, 1990.
7. Nicolaas G. de Bruijn, *Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem*, Indagationes Math. **34** (1972), 381–392.
8. Gerhard Gentzen, *Untersuchungen über das logische Schließen*, Mathematische Zeitschrift **39** (1934), 176–210, 405–431.
9. Gerhard Gentzen, *Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie*, Mathematische Annalen **119** (1943), 140–161.
10. Kurt Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatshefte für Mathematik und Physik **37** (1930), 349–360.
11. Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatshefte für Mathematik und Physik **38** (1931), 173–198.
12. David Hilbert and Paul Bernays, *Grundlagen der Mathematik II*, second ed., Grundlehren der mathematischen Wissenschaften, vol. 50, Springer–Verlag, Berlin, 1970.
13. Felix Joachimski and Ralph Matthes, *Short proofs of normalisation for the simply-typed λ-calculus, permutative conversions and Gödel's T*, Archive for Mathematical Logic **42** (2003), 59–87.
14. Ingebrigt Johansson, *Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus*, Compositio Mathematica **4** (1937), 119–136.
15. Stephen C. Kleene, *Introduction to metamathematics*, D. van Nostrand Comp., New York, 1952.
16. L. Löwenheim, *Über Möglichkeiten im Relativkalkül*, Mathematische Annalen **76** (1915), 447–470.
17. A. Malzew, *Untersuchungen aus dem Gebiete der mathematischen Logik*, Rec. Math. N. S. **1** (1936), 323–336.
18. M.H.A. Newman, *On theories with a combinatorial definition of "equivalence"*, Annals of Mathematics **43** (1942), no. 2, 223–243.
19. V.P. Orevkov, *Lower bounds for increasing complexity of derivations after cut elimination*, Zapiski Nauchnykh Seminarov Leningradskogo **88** (1979), 137–161.
20. Dag Prawitz, *Natural deduction*, Acta Universitatis Stockholmiensis. Stockholm Studies in Philosophy, vol. 3, Almqvist & Wiksell, Stockholm, 1965.
21. Kurt Schütte, *Kennzeichnung von Ordinalzahlen durch rekursiv definierte Funktionen*, Mathematische Annalen **127** (1954), 16–32.
22. J.C. Shepherdson and H.E. Sturgis, *Computability of recursive functions*, J. Ass. Computing Machinery **10** (1963), 217–255.

23. Joseph R. Shoenfield, *Mathematical logic*, Addison–Wesley Publ. Comp., Reading, Massachusetts, 1967.
24. T. Skolem, *Logisch–kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen*, Skrifter utgitt av Videnkapsselskapet i Kristiania, I, Mat. Naturv. Kl. **4** (1920), 36 pp.
25. Richard Statman, *Bounds for proof-search and speed-up in the predicate calculus*, Annals of Mathematical Logic **15** (1978), 225–287.
26. William W. Tait, *Infinitely long terms of transfinite type I*, Formal Systems and Recursive Functions (J. Crossley and M. Dummett, eds.), North–Holland, Amsterdam, 1965, pp. 176–185.
27. Anne S. Troelstra and Helmut Schwichtenberg, *Basic proof theory*, 2nd ed., Cambridge University Press, 2000.
28. Femke van Raamsdonk and Paula Severi, *On normalisation*, Computer Science Report CS-R9545 1995, Centrum voor Wiskunde en Informatica, 1995, Forms a part of van Raamsdonk's thesis from 1996.
29. Oswald Veblen, *Continuous increasing functions of finite and transfinite ordinals*, Transactions AMS **9** (1908), 280–292.

Index